AD_____

Award Number: W81XWH-12-1-0122

TITLE: Superior Volumetric Modulated Arc Therapy Planning Solution for Prostate Patients

PRINCIPAL INVESTIGATOR: Ran Davidi

CONTRACTING ORGANIZATION: Stanford University
Stanford, CA 94305-5847

REPORT DATE: April 2013

TYPE OF REPORT: Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
April 2013	Annual Summary	1 April 2012 – 31 March 2013

4. TITLE AND SUBTITLE

Superior Volumetric Modulated Arc Therapy Planning Solution for Prostate Patients

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-12-1-0122

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Ran Davidi

E-Mail: rdavidi@stanford.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Stanford University
Stanford, CA 94305-5847

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Inverse planning is at the heart of prostate Volumetric Modulated Arc Therapy (VMAT) treatment procedure and critically determines its level of success. As practiced now, the capacity of VMAT is greatly underutilized because of inferior computing performance of existing optimization methods. An alternative mathematical approach that improves both the efficiency and the efficacy is needed and is the center of this research. We propose to develop a new innovative inverse planning tool, based on the novel idea of superiorization, to replace the classical constrained optimization approaches employed in clinics today for prostate VMAT cases. Year 1 of the training award focused on formulating the VMAT problem as a constrained superiorization problem and on the development of a framework of fast converging inverse planning algorithms. Empty solution sets and infeasibility constraints that often exist in real-world applications were incorporated into the model. The new framework was proven mathematically and its efficacy was demonstrated when it was compared to a classical optimization method. The superiorization methodology was Implemented, tested and evaluated on a previously treated prostate case where good initial results were obtained.

15. SUBJECT TERMS
PROSTATE CANCER, VMAT, SUPERIORIZATION, INVERSE-PLANNING, PROJECTION-METHODS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	50	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

Introduction

Inverse planning is at the heart of prostate Volumetric Modulated Arc Therapy (VMAT) treatment procedure and critically determines its level of success. As practiced now, the capacity of VMAT is greatly underutilized because of inferior computing performance of existing optimization methods. An alternative mathematical approach that improves both the efficiency and the efficacy is needed and is the center of this research. We propose to develop a new innovative inverse planning tool, based on the novel idea of *superiorization*, to replace the classical constrained optimization approaches employed in clinics today for prostate VMAT cases.

Towards this goal, year 1 of the training award focused on formulating the VMAT problem as a constrained superiorization problem and on the development of a framework of fast converging inverse planning algorithms. The new approach was then implemented, tested and evaluated on a previously treated prostate cancer case where initial results were obtained.

Body

1. *Research Accomplishments*

SOW Aim 1: Develop algorithms for inverse planning using superiorization techniques for prostate VMAT

Meeting the goal outlined in the SOW aim 1, we have studied the problem of inverse planning for prostate VMAT and developed a framework of algorithms using the superiorization methodology that is specifically tailored to this application. We first defined the problem mathematically by reformulating it as a linear feasibility problem (instead of a minimization problem) and suggested a solution to solve it using the superiorization methodology. In developing the tools, we have also generalized the approach to include other medical physics applications, and provided conditions that are simple to meet both in theory and in practice. Our claims were proved mathematically and the results were submitted two journal (archival) publications [2, 4].

Task 1: Formulating the VMAT treatment planning as a constrained superiorization problem

Our approach to a VMAT treatment planning started by studying the current mathematical models used for this application. Since the superiorization methodology requires a different mathematical formulation, the first step was to model the problem accordingly.

Consider the system of equations

$$Ax = d, \tag{1}$$

where A is the $J \times I$ dose matrix that maps any intensity of beamlets vector $x = (x_i)_{i=1}^{I} \in R^I$ onto a dose in voxels vector $d = (d_j)_{j=1}^{J} \in R^J$. Here I is the total number of beamlets and J is the total number of voxels. The minimization problem can then be formulated as

$$\begin{array}{ll} \text{minimize} & \sum_{s=1}^{S} \lambda_s \left\| A_s x - d_{(s)} \right\|^2 \\ \text{subject to} & x \geq 0, \end{array} \tag{2}$$

where the index s stands for different structures, A_s is the submatrix of A related to structure s and $d_{(s)}$ is the subvector of d related to structure s, respectively, and λ_s is the importance factor associated with the

sth structure which is decided by the planner. There is an assumption that x is achievable using apertures (aperture constraints).

Assume that we have S structures, for $s = 1, 2, \ldots, S$, (including the complement of all identified structures), and denote by O_s the set of indices of voxels that belong to the sth structure, such that

$$O_s = \{j_{s,1}, j_{s,2}, \ldots j_{s,m(s)}\} \tag{3}$$

where $m(s)$ is the number of voxels in this structure. Then the system matrix A can be partitioned into blocks

$$A = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_S \end{bmatrix} \tag{4}$$

so that a submatrix A_s will contain the rows from A whose indices appear in O_s, (similarly, let $d_{(s)}$ be the subvector of d whose component indices appear in O_s) and then the system becomes

$$\begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_S \end{bmatrix} x = \begin{pmatrix} d_{(1)} \\ d_{(2)} \\ \vdots \\ d_{(S)} \end{pmatrix}. \tag{5}$$

An optimization method aims at satisfying the system (1) (equivalently (5)) while minimizing a given objective function.

Reformulating the problem as a constrained superiorization problem: We suggest the following modifications to the above modality. Replace the prescription method that gives rise to the system $Ax = d$ in (1) by a more flexible one in which we ask the planner to provide lower- and upper- dose bounds vectors, $\underline{d}$ and $\overline{d}$, respectively, on all voxels in all structures, and instead of (2) we aim at solving the following linear feasibility problem

$$\underline{d} \leq Ax \leq \overline{d}. \tag{6}$$

By transforming the problem of (1) into a linear feasibility problem of the form (6), we allow many iterative projection method to derive a solution. This enables a formulation for the superiorization methodology to be applied to VMAT inverse planning problem since many of these algorithms are also perturbation resilient. Specifically, methods that belong to the two classes of projection methods, String Averaging Projection (SAP) and Block-Iterative Projection (BIP) methods, can be applied towards solving this formulation and achieve finding a superior solution in addition to satisfying the feasibility constraints (see [1, 3]). That is, an x obtainable by a projection method alone will be an intensity of beamlets vector trying to solve (6), while using a projection method that is also perturbation resilient allows for obtaining an x that solves (6) but also provides a solution that is *superior* with respect to an objective function.

The solution vector x of the beamlet intensities that results from the superiorization approach will then be evaluated. Tools such as dose volume histograms (DVHs) will help assess conformality to the prostate (the target) and to the organs at risk (OAR). These will be compared against what is recommended by a physician in the clinic and governed by the specifications of the Radiation Treatment Oncology Group (RTOG) protocol for prostate cancer patients [5].

The adaptation to our model based on the RTOG protocol is as follows: Given a structure s that is an OAR, we define $\overline{d_{(s)}}$ to be the upper-bound subvector of the prescribed dose

$$\overline{d_{(s)}} \equiv d_{(s)}, \tag{7}$$

and define $\underline{d_{(s)}}$ to be a lower-bound subvector for any target structure s

$$\underline{d_{(s)}} \equiv d_{(s)}. \tag{8}$$

This allows the constraints in (6) to be written as

$$0 \leq A_s x \leq d_{(s)}, \tag{9}$$

for an OAR structure s and as

$$d_{(s)} \leq A_s x \leq e_{(s)}, \tag{10}$$

for a target structure s, where $e_{(s)}$ is a clinically-specified upper-bound subvector for the target.

In assessing the solution provided by the superiorization method, if the acceptance criteria is not met, then a refined selection of $\underline{d}$ and $\overline{d}$ will be provided and the process will repeat until a superior feasible solution is found (this step is identical to how it is done in the clinic today).

Task 2: Development of a framework for fast converging inverse planning superiorization techniques
Task 7: Investigate the underlying principles and put their concept on a firm mathematical ground

In developing a framework for fast converging inverse planning superiorization techniques we first identified several problems that currently exist in optimization methods. In classical optimization it is assumed that there is a constraints set C and the task is to find an $x \in C$ for which $\phi(x)$ is minimal. Problems with this approach are the following: (1) The constraints may not be consistent and so C could be empty and the optimization task as stated would not have a solution. (2) Iterative methods of classical constrained optimization typically converge to a solution only in the limit. In practice some stopping rule is applied to terminate the process and the actual output at that time may not be in C and, even if it is in C, it is most unlikely to be a minimizer of ϕ over C.

Both problems were addressed in the newly developed superiorization framework. Mathematical definitions and conditions were introduced and were theoretically proven. The new foundations include two new notions of *constraints-compatibility* and *strong perturbation resiliency*. The new concepts allow to take into the modality the infeasibility and practical convergence problems that exist in optimization methods. More specifically, in the superiorization model we suggested to replace the constraints set C by a nonnegative real-valued function Pr that serves as an indicator of how incompatible a given x is with the constraints. Then the merit of an actual output of an algorithm is given by the smallness of the two numbers $Pr(x)$ and $\phi(x)$. Roughly, if an iterative algorithm produces an output x, then its superiorized version will produce an output x' for which $Pr(x')$ is not larger then $Pr(x)$, but (in general) $\phi(x')$ is much smaller than $\phi(x)$.

In addition to the theoretical developments of superiorization, a practical and systematic way was developed to turn any iterative algorithm that solves a feasibility problem into an algorithm that does superiorization. For an iterative algorithm **P** and for any optimization criterion ϕ for which we know how to produce nonascending vectors (see definition p. 5536 in [4]), the following pseudocode automatically takes **P** and produces a version of **P** that is superiorized for ϕ (exact details of the procedure can be found on page 5537 in [4]):

Superiorized Version of the Basic Algorithm

1. **set** $k = 0$
2. **set** $y^k = y^0$
3. **set** $\ell = -1$
4. **repeat**
5. **set** $n = 0$
6. **set** $y^{k,n} = y^k$
7. **while** $n<N$
8. **set** $v^{k,n}$ to be a nonascending vector for ϕ at $y^{k,n}$
9. **set** *loop=true*
10. **while** *loop*
11. **set** $\ell = \ell + 1$
12. **set** $\beta_{k,n} = \eta_\ell$
13. **set** $z = y^{k,n} + \beta_{k,n} v^{k,n}$
14. **if** $\phi(z) \leq \phi\left(y^k\right)$ **then**
15. **set** $n = n + 1$
16. **set** $y^{k,n} = z$
17. **set** *loop = false*
18. **set** $y^{k+1} = \boldsymbol{A}_C\left(y^{k,N}\right)$
19. **set** $k = k + 1$

By bridging the gap that typically exist between theory and practice in the new model, superiorization was made more general. That is, the framework fit many other medical physics application, not just VMAT or radiation therapy inverse planning type applications. All the results mentioned briefly here have been published in an archival journal publication in the journal of Medical Physics [4], see the Appendix Section for the full manuscript.

Another accomplishment related to this task touches on an additional aspect of superiorization. Constrained optimization problems that arise in real-life applications are often huge (such an example is the VMAT problem). It can then happen that the traditional algorithms for constrained optimization require computational resources that are not easily available and, even if they are, the length of time needed to produce an acceptable output is too long to be practicable. As part of our goal to show that superiorization can handle large size problems efficiently, we have illustrated that the computational requirements of a superiorized algorithm can be significantly less than that of a traditional optimization algorithm, by reporting on a comparison of superiorization with the projected subgradient method (PSM), which is a standard

	$Total\,Variation$ value	Time (seconds)
projected subgradient method	919	2217
superiorization method	873	102

Table 1: Performance comparison of the projected subgradient method and the superiorization method with Total Variation as the objective function.

method of classical optimization. Table 1 summarizes the comparison we have performed between the PSM method and the superiorization method. In our experiment, we set the the stopping rule to guarantee that the output of the superiorization method is at least as constraints-compatible as the output of the PSM. The superiorization method showed clearly superior efficacy to the PSM: it obtained a result with a lower objective function value (TV) at less than one twentieth of the computational cost.

The complete report that summarizes this work was submitted to the Journal of Optimization Theory and Applications and is currently under review [2]. It is attached to this report in the Appendix Section.

Task 3: Implementation and testing of the developed algorithms

In this task we wanted to assess our proposed approach to using superiorization on a realistically yet simple test case. The goal set here is two-fold: the first is to show that the developed method can produce good results and the second is to obtain a clear indication if the nonacsending-type superiorization techniques should be replaced with alternative derivative-free approaches (see, SOW Task 4: the development of alternative derivative-free techniques to superiorization).

We proposed to use as a test case in this task a previously treated intensity modulated radiation therapy (IMRT) prostate patient case. As was explained in the research proposal, the VMAT technique delivers an IMRT type treatment in a single arc. Getting good results on a previously treated IMRT case would establish a level of confidence that the superiorization method can deliver superior results by referencing a previously treated clinical case. The modality that was given above (in Task 1) is identical for these two radiation therapy techniques (i.e., IMRT and VMAT); the difference lies in the size of the problem and its level of complexity. Since superiorization was never tried with any type of radiation therapy treatment planning it is important to provide such evidence on an actual clinical case. The mathematical model which we have developed along with the theories and proofs of the superiorization methodology (in Task 2) fit both problems. Satisfactory results will encourage us to continue develop the method as it is proposed in Tasks 1 and 2 and tailor it further more to the VMAT approach.

Algorithmic operator and objective function The framework that was developed is quite general for many medical physics applications. With the modality of the superiorization approach in (6), a choice for a projection operator that is perturbation resilient is needed as well as a choice of an objective function. The algorithmic operator that was chosen for our implementation was the Algebraic Reconstruction Technique (ART) for inequalities constraints. This operator was proven to be perturbation-resilient in [1]. The constraints of the system in (6) can be thought of as hyperslabs. The algorithm projects the current point according to its location in relation to the two hyperplanes that form a hyperslab. A geometrical description to this feasibility problem is provided in Figure 1. The analytical formulation associated with it is the

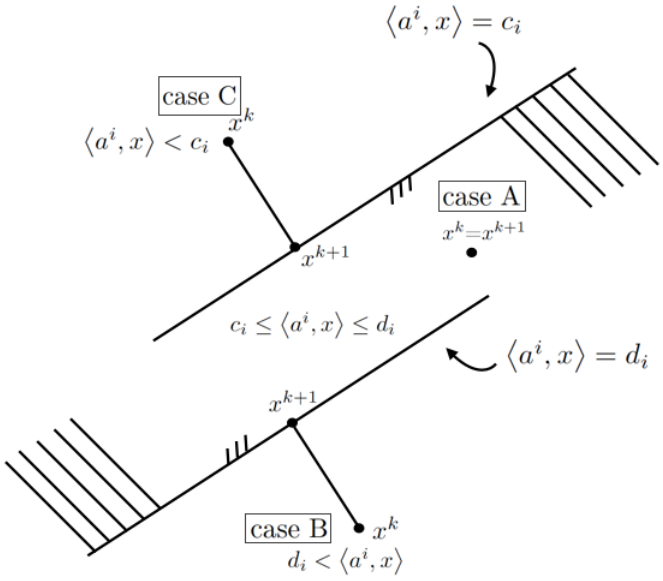


Figure 1: Geometrical description of ART with inequalities constraints

following:

$$x^{k+1} = \begin{cases} x^k, & \text{if } c_i \leq \langle a^i, x^k\rangle \leq d_i \text{ (case A)}, \\ x^k + \lambda_k \frac{d_i - \langle a^i, x^k\rangle}{\|a^i\|^2} a^i, & \text{if } d_i < \langle a^i, x^k\rangle \text{ (case B)}, \\ x^k + \lambda_k \frac{c_i - \langle a^i, x^k\rangle}{\|a^i\|^2} a^i, & \text{if } \langle a^i, x^k\rangle < c_i \text{ (case C)}. \end{cases} \tag{11}$$

The objective function used in our implementation was the total variation (TV) functional of the beamlet intensity vector x, see Eq. (12) in [4] and the discussion in the research proposal under Specific Aim 1 regarding this choice. We denote herein the superiorization algorithm that uses TV as the objective function by TV-Superiorization.

Prostate patient data and planning The data for testing the approach were of a previously treated prostate cancer patient. A seven field radiation treatment IMRT plan was created. The organs that were included in the plan were the prostate (target), rectum, bladder, small bowel (OARs) and the full body. Figure 2 shows the CT and the contours of these structures. Using RTOG 0815 [5] we set in Table 2 the acceptance criteria for the implemented TV-Superiorization algorithm.

Results We compared the results when superiorization was present versus when it was not. The TV-Superiorization algorithm was able to meet the RTOG acceptance criteria while the one without TV-Superiorization was not. Moreover, the TV-Superiorization algorithm was able to achieve this in a relatively short amount time of only 12 iterations. Figure 3 shows the DVH curves of the two algorithms side-by-side.

Figure 2: CT of the prostate patient case used in the experiment.

Organ	Target?	Acceptance criteria
I. Prostate	Yes	1. Dose will be normalized s.t. 98% of the PTV receives the prescription dose. (Prescribed dose to PTV is 79.2 Gy.) 2. The maximum allowable dose within the PTV is 107% of the prescribed dose (i.e., maximum allowed dose is 84.744 Gy). 3. The minimum allowable dose within the PTV should be $\geq$95% of the prescribed dose (i.e., 100% of the dose should be $\geq$75.24 Gy.
II. Rectum	No	1. No more than 15% volume receives dose that exceeds 75 Gy 2. No more than 25% volume receives dose that exceeds 70 Gy 3. No more than 35% volume receives dose that exceeds 65 Gy 4. No more than 50% volume receives dose that exceeds 60 Gy
III. Bladder	No	1. No more than 15% volume receives dose that exceeds 80 Gy 2. No more than 25% volume receives dose that exceeds 75 Gy 3. No more than 35% volume receives dose that exceeds 70 Gy 4. No more than 50% volume receives dose that exceeds 65 Gy
IV. Small Bowel	No	1. Upper bound is set to 52 Gy.

Table 2: Acceptance criteria for prostate patients according to RTOG 0815 [5].

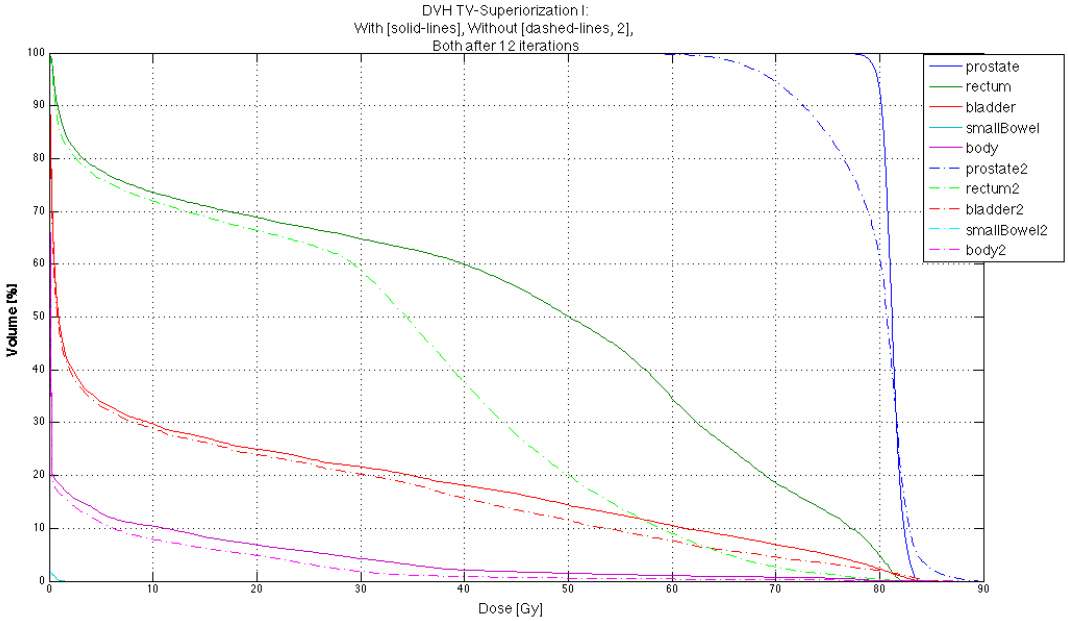


Figure 3: DVH plots for a prostate case experiment with and without TV-Superiorization.

The solid lines represent the TV-Superiorization algorithm and the dashed lines represent the algorithm without Superiorization. The corresponding numbers for assessing the acceptance criteria are specified in Table 3. As can be seen, the criteria that is based on the RTOG protocol [5] was fully met by the superiorization method for this prostate case.

Training Accomplishments

Task 8: Seminar, lectures and meetings

Task 9: Research training

Task 10: Clinical training

During year 1 of the training award the PI had attended regular meetings, seminars and journal clubs with presentations on topics related to radiation therapy treatment planning. Other presentations of visiting scholars and professionals were also available throughout the year and had enriched his knowledge on the topic. The PI was trained in the clinic to operate the Eclipse system stations for treatment planning available at Stanford Cancer Center (Eclipse is a commercial tool for treatment planning developed by Varian Medical Systems). He collaborated with radiation oncologists, radiation therapists, physicists and dosimetrists and obtained first-hand the knowledge and experience of the process of prostate radiation treatment planning.

Organ	Target?	Criterion	TV-Superiorization
I. Prostate	Yes	%vol $>$ 79.2 Gy $=$ 98	%vol $>$ 79.2 Gy $=$ 98
		%vol $>$ 84.744 Gy $=$ 0	%vol $>$ 84.6 Gy $=$ 0
		%vol $>$ 75.24 Gy $=$ 100	%vol $>$ 75.24 Gy $=$ 100
II. Rectum	No	%vol $>$ 75 Gy $\leq$ 15	%vol $>$ 75 Gy $\leq$ 12.7
		%vol $>$ 70 Gy $\leq$ 25	%vol $>$ 70 Gy $\leq$ 18.6
		%vol $>$ 65 Gy $\leq$ 35	%vol $>$ 65 Gy $\leq$ 25.8
		%vol $>$ 60 Gy $\leq$ 50	%vol $>$ 60 Gy $\leq$ 34.5
III. Bladder	No	%vol $>$ 80 Gy $\leq$ 15	%vol $>$ 80 Gy $\leq$ 2.2
		%vol $>$ 75 Gy $\leq$ 25	%vol $>$ 75 Gy $\leq$ 4.9
		%vol $>$ 70 Gy $\leq$ 35	%vol $>$ 70 Gy $\leq$ 6.8
		%vol $>$ 65 Gy $\leq$ 50	%vol $>$ 65 Gy $\leq$ 8.7
IV. Small Bowel	No	%vol $>$ 52 Gy $\leq$ 0	%vol $>$ 1.4 Gy $\leq$ 0

Table 3: Results of the criteria for the TV-Superiorization algorithm.

Key Research Accomplishments

- Formulated the VMAT treatment planning as a constrained superiorization problem.
- Developed a framework for fast converging inverse planning superiorization techniques.
- Derived the necessary conditions of the superiorization framework for VMAT treatment planning
- Placed the newly developed concepts on a firm mathematical ground.
- Implemented and tested the new superiorization framework and showed good initial results.
- Trained in treating prostate cancer as it is done in the clinic today.

Reportable Outcomes

- Two journal publications were submitted. One appeared in the journal of Medical Physics and another is currently under review:

  1. G.T. Herman, E. Garduño, R. Davidi and Y. Censor, Superiorization: An optimization heuristic for medical physics, *Medical Physics* **39** (2012), 5532–5546.
  2. Y. Censor, R. Davidi, G.T. Herman, R.W. Schulte and L. Tetruashvili, Projected subgradient minimization versus superiorization, Journal of Optimization Theory and Applications (submitted).

- The above work has been accepted for presentation at the joint workshop sponsored by the American Society for Therapeutic Radiology and Oncology (ASTRO), the National Cancer Institute (NCI) and the American Association of Physicists in Medicine (AAPM) , June 13-14, 2013, National Institutes of Health, Bethesda, MD, USA.

- The above work has been accepted for presentation at the workshop on Projection Methods: Theory and Practice, June 19-21, 2013, Fraunhofer Institute for Industrial Mathematics ITWM, Kaiserslautern, Germany.

Conclusion

In year 1 we were able to extend the superiorization methodology into a larger framework, one that is more realistic from the point of view of the application at hand. by taking into account the discrepancy that exist between theory and practice and incorporate it into our model, we minimized potential issues that typically appear when a theory is applied to a real-life application.

Superiorization was developed to be a general tool for medical physics applications. It is capable of turning any iterative algorithm that tries to satisfy a set of constraints into one that is also capable of superiorizing an objective function. The work that came out of this research can help other applications that use optimization methods as the main tool.

Using the above methodology, we tailored it specifically to solve the problem of VMAT in radiation therapy inverse planning. The initial results obtained on a realistic IMRT prostate case were satisfactory and show good indication that superiorization works and can be applied to a radiation treatment planning problem such as IMRT and VMAT.

Our next steps include the full implementation, testing and evaluation of VMAT prostate cases. A thorough investigation as detailed in the SOW is planned in year 2 of this training award.

References

[1] D. Butnariu, R. Davidi, G.T. Herman and I.G. Kazantsev, Stable convergence behavior under summable perturbations of a class of projection methods for convex feasibility and optimization problems, *IEEE Journal of Selected Topics in Signal Processing* **1** (2007), 540–547.

[2] Y. Censor, R. Davidi, G.T. Herman, R.W. Schulte and L. Tetruashvili, Projected subgradient minimization versus superiorization, Journal of Optimization Theory and Applications (submitted).

[3] R. Davidi, G.T. Herman and Y. Censor, Perturbation-resilient block-iterative projection methods with application to image reconstruction from projections, *International Transactions in Operational Research* **16** (2009), 505–524.

[4] G.T. Herman, E. Garduño, R. Davidi and Y. Censor, Superiorization: An optimization heuristic for medical physics, *Medical Physics* **39** (2012), 5532–5546.

[5] A. Martinez *et al.* A phase III prospective randomized trial of dose-escalated radiotherapy with or without short-term androgen deprivation therapy for patients with intermediate-risk prostate cancer. Last accessed April 19, 2013. http://www.rtog.org/clinicaltrials/protocoltable/studydetails.aspx?action=openFile&FileID=4662

Appendices

Superiorization: An optimization heuristic for medical physics

Gabor T. Herman[a)]
Department of Computer Science, The Graduate Center, City University of New York, New York, New York 10016

Edgar Garduño
Departamento de Ciencias de la Computación, Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autónoma de México, Cd. Universitaria, Mexico City C.P. 04510, Mexico

Ran Davidi
Department of Radiation Oncology, Stanford University, Stanford, California 94305

Yair Censor
Department of Mathematics, University of Haifa, Mt. Carmel, 31905 Haifa, Israel

(Received 12 January 2012; revised 28 July 2012; accepted for publication 29 July 2012; published 22 August 2012)

Purpose: To describe and mathematically validate the superiorization methodology, which is a recently developed heuristic approach to optimization, and to discuss its applicability to medical physics problem formulations that specify the desired solution (of physically given or otherwise obtained constraints) by an optimization criterion.
Methods: The superiorization methodology is presented as a heuristic solver for a large class of constrained optimization problems. The constraints come from the desire to produce a solution that is constraints-compatible, in the sense of meeting requirements provided by physically or otherwise obtained constraints. The underlying idea is that many iterative algorithms for finding such a solution are perturbation resilient in the sense that, even if certain kinds of changes are made at the end of each iterative step, the algorithm still produces a constraints-compatible solution. This property is exploited by using permitted changes to steer the algorithm to a solution that is not only constraints-compatible, but is also desirable according to a specified optimization criterion. The approach is very general, it is applicable to many iterative procedures and optimization criteria used in medical physics.
Results: The main practical contribution is a procedure for automatically producing from any given iterative algorithm its superiorized version, which will supply solutions that are superior according to a given optimization criterion. It is shown that if the original iterative algorithm satisfies certain mathematical conditions, then the output of its superiorized version is guaranteed to be as constraints-compatible as the output of the original algorithm, but it is superior to the latter according to the optimization criterion. This intuitive description is made precise in the paper and the stated claims are rigorously proved. Superiorization is illustrated on simulated computerized tomography data of a head cross section and, in spite of its generality, superiorization is shown to be competitive to an optimization algorithm that is specifically designed to minimize total variation.
Conclusions: The range of applicability of superiorization to constrained optimization problems is very large. Its major utility is in the automatic nature of producing a superiorization algorithm from an algorithm aimed at only constraints-compatibility; while nonheuristic (exact) approaches need to be redesigned for a new optimization criterion. Thus superiorization provides a quick route to algorithms for the practical solution of constrained optimization problems. © *2012 American Association of Physicists in Medicine*. [http://dx.doi.org/10.1118/1.4745566]

Key words: superiorization, constrained optimization, heuristic optimization, tomography, total variation

I. INTRODUCTION

Optimization is a tool that is used in many areas of Medical Physics. Prime examples are radiation therapy treatment planning and tomographic reconstruction, but there are others such as image registration. Some well-cited classical publications on the topic are Refs. 1–12 and some recent articles are Refs. 13–26.

In a typical medical physics application, one uses *constrained optimization*, where the constraints come from the desire to produce a solution that is *constraints-compatible*, in the sense of meeting the requirements provided by physically or otherwise obtained constraints. In radiation therapy treatment planning, the requirements are usually in the form of constraints prescribed by the treatment planner on the doses to be delivered at specific locations in the body. These doses in turn depend on information provided by an imaging instrument, typically a magnetic resonance imaging (MRI) or a computerized tomography (CT) scanner. In tomography, the constraints come from the detector readings of the instrument.

In such applications, it is typically the case that a large number of solutions would be considered good enough from the point of view of being constraints-compatible; to a large extent, but not entirely, due to the fact that there is uncertainty as to the exact nature of the constraints (for example, due to noise in the data collection). In such a case, an optimization criterion is introduced that helps us to distinguish the "better" constraints-compatible solutions (for example, this criterion could be the total dose to be delivered to the body, which may vary quite a bit between radiation therapy treatment plans that are compatible with the constraints on the doses delivered to individual locations).

The superiorization methodology (see, for example, Refs. 22 and 27–32) is a recently developed heuristic approach to optimization. The word *heuristic* is used here in the sense that the process is not guaranteed to lead to an optimum according to the given criterion; approaches aimed at processes that are guaranteed in that sense are usually referred to as *exact*. Heuristic approaches have been found useful in practical applications of optimization, mainly because they are often computationally much less expensive than their exact counterparts, but nevertheless provide solutions that are appropriate for the application at hand.[33–35]

The underlying idea of the superiorization approach is the following. In many applications there exists a computationally efficient iterative algorithm that produces a constraints-compatible solution for the given constraints. (An example of this for radiation therapy treatment planning is reported in Ref. 36, its clinical use is discussed in Ref. 15.) Furthermore, often the algorithm is *perturbation resilient* in the sense that, even if certain kinds of changes are made at the end of each iterative step, the algorithm still produces a constraints-compatible solution.[27–30] This property is exploited in the superiorization approach by using such perturbations to steer the algorithm to a solution that is not only constraints-compatible, but is also desirable according to a specified optimization criterion. The approach is very general, it is applicable to many iterative procedures and optimization criteria.

The current paper presents a major advance in the practice and theory of superiorization. The previous publications[22,27–32] used the intuitive idea to present some superiorization algorithms, in this paper the reader will find a totally automatic procedure that turns an iterative algorithm into its superiorized version. This version will produce an output that is as constraints-compatible as the output of the original algorithm, but it is superior to that according to an optimization criterion. This claim is mathematically shown to be true for a very large class of iterative algorithms and for optimization criteria in general, typical restrictions (such as convexity) on the optimization criterion are not essential for the material presented below. In order to make precise and validate this broad claim, we present here a new theoretical framework. The framework of Ref. 29 is a precursor of what we present here, but it is a restricted one, since it assumes that the constraints can be all satisfied simultaneously, which is often false in medical physics applications. There is no such restriction in the presentation below.

The idea of designing algorithms that use interlacing steps of two different kinds (in our case, one kind of steps aim at constraints-compatibility and the other kind of steps aim at improvement of the optimization criterion) is well-established and, in fact, is made use of in many approaches that have been proposed with exact constrained optimization in mind; see, for example, the works of Helou Neto and De Pierro,[37,38] Nurminski,[39] Combettes and co-workers,[40,41] Sidky and co-workers,[23,42,43] and Defrise and co-workers.[44] However, none of these approaches can do what can be done by the superiorization approach as presented below, namely, the automatic production of a heuristic constrained optimization algorithm from an iterative algorithm for constraints-compatibility. For example, in Ref. 37 it is assumed (just as in the theory presented in our Ref. 29) that all the constraints can be satisfied simultaneously.

A major motivator for the additional theory presented in the current paper is to get rid of this assumption, which is not reasonable when handling real problems of medical physics. Motivated by similar considerations, Helou Neto and De Pierro[38] present an alternative approach that does not require this unreasonable assumption. However, in order to solve such a problem, they end up with iterative algorithms of a particular form rather than having the generality of being able to turn any constraints-compatibility seeking algorithm into a superiorized one capable of handling constrained optimization. Also, the assumptions they have to make in order to prove their convergence result (their Theorem 15) indicate that their approach is applicable to a smaller class of constrained optimization problems than the superiorization approach whose applicability seems to be more general. However, for the mathematical purist, we point out that they present an exact constrained optimization algorithm, while superiorization is a heuristic approach. Whether this is relevant to medical physics practice is not clear: exact algorithms are not run forever, but are stopped according to some stopping-rule, the relevant questions in comparing two algorithms are the quality of the actual output and the computation time needed to obtain it.

Ultimately, the quality of the outputs should be evaluated by some figures of merit relevant to the medical task at hand. An example of a careful study of this kind that involves superiorization is in Sec. 4.3 of Ref. 30, which reports on comparing in CT the efficacy of constrained optimization reconstruction algorithms for the detection of low-contrast brain tumors by using the method of statistical hypothesis testing (which provides a P-value that indicates the significance by which we can reject the null hypothesis that the two algorithms are equally efficacious in favor of the alternative that one is preferable). Such studies bundle together two things: (i) the formulation of the constrained optimization task and (ii) the performance of the algorithm in performing that task. The first of these requires a translation of the medical aim into a mathematical model, it is important that this model should be appropriately chosen.

The superiorization approach is not about choosing this model, it kicks in once the model is chosen and aims at producing an output that is "good" according to the

mathematical specifications of the constraints and of the optimization criterion. Thus superiorization has been used to compare the effects on the quality of the output in CT when the optimization criterion is specified by total variation (TV) versus by entropy[28] or versus by the ℓ_1-norm of the Haar transform.[32] However, the current paper is not about discussing how to translate the underlying medical physics task into a constrained optimization problem. For our purposes here, we are assuming that the mathematical model has been worked out and concentrate on the algorithmic approach for solving the resulting constrained optimization problem. We claim that the evaluation of such algorithms should not be based on the medical figures of merit mentioned at the beginning of the previous paragraph, but rather on their performance in solving the mathematical problem. If "good" solutions to the constrained optimization problem are not medically efficacious, that indicates that something is wrong with the mathematical model and not that something is wrong with the algorithmic approach. For this reason, in this paper we will not carry out a careful investigation of the medical efficacy of any algorithm in the manner that we have done in Sec. 4.3 of Ref. 30, but will restrict ourselves to a simple illustration of the performance of the superiorization approach as compared to the previously published algorithm of Ref. 42 that is aimed at performing exact minimization.

Examples of such studies already exist. Superiorization was compared in Ref. 27 with Algorithm 6 of Ref. 40 and in Ref. 45 with the algorithm of Goldstein and Osher that they refer to as TwIST (Ref. 46) with split Bregman[47] as the substep. In both cases the implementation was done by the proposers of the algorithms. In these reported instances superiorization did well: the constraints-compatibility and the value of the function to be minimized were very similar for the outputs produced by the algorithms being compared, but the superiorization algorithm produced its output four times faster than the alternative. It would be unjustified to draw any general conclusions on the mathematical performance and speed of superiorization based on just a few experiments, but the reported results are encouraging.

However, the main reason why we advocate superiorization is different from what is discussed above. The reason why we claim it to be helpful in medical physics research is that it has the potential of saving a lot of time and effort for the researcher. Let us consider a historical example. Likelihood optimization using the iterative process of expectation maximization (EM) (Ref. 48) gained immediate and wide acceptance in the emission tomography community. It was observed that irregular high amplitude patterns occurred in the image with a large number of iterations, but it was not until five years later that this problem was corrected[49] by the use of a maximum a posteriority probability (MAP) algorithm with a multivariate Gaussian prior. Had we had at our disposal the superiorization approach, then the introduction of an optimization criterion (Gaussian or other) into the iterative EM process would have been a simple matter and we would have saved the time and effort spent on designing a special purpose algorithm for the MAP formulation. A TV-superiorization of the EM algorithm is presented in Ref. 50.

Even though our major claim for superiorization is that it provides a quick route to algorithms for the practical solution of constrained optimization problems, before leaving this introduction let us bring up a question that has to do with the performance of the resulting algorithms: Will superiorization produce superior results to those produced by contemporary MAP methods or is it faster than the better of such methods? At this stage we have not yet developed the mathematical notation to discuss this question in a rigorous manner, we return to it in Subsection II.F.

In Sec. III, we present in detail the superiorization methodology. In Sec. III, we provide an illustrative example by reporting on reconstructions produced by algorithms applied to simulated computerized tomography data of a head cross section. In the final section, we discuss our results and present our conclusions.

II. THE SUPERIORIZATION METHODOLOGY

II.A. Problem sets, proximity functions, and ε-compatibility

Although optimization is often studied in a more general context (such as in Hilbert or Banach spaces), in medical physics we usually deal with a special case, where optimization is performed in a *Euclidean space* $\mathbb{R}^J$ (the space of J-dimensional vectors of real numbers, where J is a positive integer). As often appropriate in practice, we further restrict the domain of optimization to a nonempty subset Ω of $\mathbb{R}^J$ (such as the *non-negative orthant* $\mathbb{R}^J_+$ that consists of vectors all of whose components are non-negative).

We now turn to formalizing the notion of being compatible with given constraints, a notion that we have used informally in Sec. I. In any application, there is a *problem set* $\mathbb{T}$; each *problem* $T \in \mathbb{T}$ is essentially a description of the constraints in that particular case. For example, for a tomographic scanner, the problem of reconstruction for a particular patient at a particular time is determined by the measurements taken by the scanner for that patient at that time. The intuitive notion of constraints-compatibility is formalized by the use of a *proximity function* $\mathcal{P}r$ on $\mathbb{T}$ such that, for every $T \in \mathbb{T}$, $\mathcal{P}r_T$ maps Ω into $\mathbb{R}_+$, the set of non-negative real numbers; i.e., $\mathcal{P}r_T : \Omega \to \mathbb{R}_+$. Intuitively, we think of $\mathcal{P}r_T(\boldsymbol{x})$ as an indicator of how incompatible $\boldsymbol{x}$ is with the constraints of T. For example, in tomography, $\mathcal{P}r_T(\boldsymbol{x})$ should indicate by how much a proposed reconstruction that is described by an $\boldsymbol{x}$ in Ω violates the constraints of the problem T that are provided by the measurements taken by the scanner. For example, if we use $\boldsymbol{b}$ to denote the vector of estimated line integrals based on the measurements obtained by the scanner and by $\boldsymbol{A}$ the system matrix of the scanner, then a possible choice for the proximity function is the norm-distance $\|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|$, which we will use as an example in the discussions that follow. An alternative legitimate choice for the proximity function is the Kullback-Leibler distance $KL(\boldsymbol{b}, \boldsymbol{A}\boldsymbol{x})$, which is the negative log-likelihood of a statistical model in tomography. The

special case $\mathcal{P}r_T(\boldsymbol{x}) = 0$ is interpreted by saying that $\boldsymbol{x}$ is perfectly compatible with the constraints; due to the presence of noise in practical applications, it is quite conceivable that there is no $\boldsymbol{x}$ that is perfectly compatible with the constraints, and we accept an $\boldsymbol{x}$ as constraints-compatible as long as the value of $\mathcal{P}r_T(\boldsymbol{x})$ is considered to be small enough to justify that decision. Combining these two concepts leads to the notion of a *problem structure,* which is a pair $\langle \mathbb{T}, \mathcal{P}r \rangle$, where $\mathbb{T}$ is a nonempty problem set and $\mathcal{P}r$ is a proximity function on $\mathbb{T}$. For a problem structure $\langle \mathbb{T}, \mathcal{P}r \rangle$, a problem $T \in \mathbb{T}$, a non-negative ε, and an $\boldsymbol{x} \in \Omega$, we say that $\boldsymbol{x}$ is *ε-compatible* with T provided that $\mathcal{P}r_T(\boldsymbol{x}) \leq \varepsilon$.

As an example (whose applicability to tomographic reconstruction is illustrated in Sec. III), consider the problem structure that arises from the desire to find non-negative solutions of sequences of blocks of linear equations. Then the appropriate choices are $\Omega = \mathbb{R}_+^J$ and the problem structure is $\langle \mathbb{S}, Res \rangle$, where the problem set $\mathbb{S}$ is

$$\begin{aligned}\mathbb{S} = \{&(\{(\boldsymbol{a}^1, b_1), \ldots, (\boldsymbol{a}^{\ell_1}, b_{\ell_1})\}, \ldots,\\ &\{(\boldsymbol{a}^{\ell_1+\ldots+\ell_{W-1}+1}, b_{\ell_1+\ldots+\ell_{W-1}+1}), \ldots, (\boldsymbol{a}^{\ell_1+\ldots+\ell_W}, b_{\ell_1+\ldots+\ell_W})\}) |\\ &W \text{ is a positive integer and,}\\ &\text{for } 1 \leq w \leq W,\ \ell_w \text{ is a positive integer and,}\\ &\text{for } 1 \leq i \leq \ell_1 + \ldots + \ell_W,\ \boldsymbol{a}^i \in \mathbb{R}^J \text{ and } b_i \in \mathbb{R}\} \end{aligned} \tag{1}$$

and the proximity function Res on $\mathbb{S}$ is defined, for any problem $S = (\{(\boldsymbol{a}^1, b_1), \ldots, (\boldsymbol{a}^{\ell_1}, b_{\ell_1})\}, \ldots, \{(\boldsymbol{a}^{\ell_1+\ldots+\ell_{W-1}+1}, b_{\ell_1+\ldots+\ell_{W-1}+1}), \ldots, (\boldsymbol{a}^{\ell_1+\ldots+\ell_W}, b_{\ell_1+\ldots+\ell_W})\})$ in $\mathbb{S}$ and for any $\boldsymbol{x} \in \Omega$, by

$$Res_S(\boldsymbol{x}) = \sqrt{\sum_{i=1}^{\ell_1+\ldots+\ell_W} (b_i - \langle \boldsymbol{a}^i, \boldsymbol{x} \rangle)^2}. \tag{2}$$

Note that each element of this problem set $\mathbb{S}$ specifies an ordered sequence of W blocks of linear equations of the form $\langle \boldsymbol{a}^i, \boldsymbol{x} \rangle = b_i$ where $\langle *, * \rangle$ denotes the inner product in $\mathbb{R}^J$ (and thus $\mathbb{S}$ is an appropriate representation of the so-called "ordered subsets" approach to tomographic reconstruction,[51] as well as of other earlier-published block-iterative methods that proposed essentially the same idea[52–54]). The proximity function Res on $\mathbb{S}$ is the *residual* that we get when a particular $\boldsymbol{x}$ is substituted into all the equations of a particular problem S.

II.B. Algorithms and outputs

We now define the concept of an algorithm in the general context of problem structures. For technical reasons that will become clear as we proceed with our development, we introduce an additional set Δ, such that $\Omega \subseteq \Delta \subseteq \mathbb{R}^J$. (Both Ω and Δ are assumed to be known and fixed for any particular problem structure $\langle \mathbb{T}, \mathcal{P}r \rangle$.) An *algorithm* $\mathbf{P}$ for a problem structure $\langle \mathbb{T}, \mathcal{P}r \rangle$ assigns to each problem $T \in \mathbb{T}$ an operator $\mathbf{P}_T : \Delta \to \Omega$. This definition is used to define iterative processes that, for any *initial point* $\boldsymbol{x} \in \Omega$, produce the (potentially) infinite sequence $((\mathbf{P}_T)^k \boldsymbol{x})_{k=0}^{\infty}$ (that is, the sequence $\boldsymbol{x}, \mathbf{P}_T\boldsymbol{x}, \mathbf{P}_T(\mathbf{P}_T\boldsymbol{x}), \cdots$) of points in Ω. We discuss below how such a potentially infinite process is terminated in practice.

Selecting $\Omega = \mathbb{R}_+^J$ and $\Delta = \mathbb{R}^J$ for the problem structure $\langle \mathbb{S}, Res \rangle$ of Subsection II.A, an example of an algorithm $\mathbf{R}$ is specified by

$$\mathbf{R}_S\boldsymbol{x} = \mathbf{Q}\mathbf{B}_{S_W} \cdots \mathbf{B}_{S_1}\boldsymbol{x}, \tag{3}$$

where S is the problem specified above in Eq. (2) and, for $1 \leq w \leq W$, $\mathbf{B}_{S_w} : \Delta \to \Delta$ is defined by

$$\mathbf{B}_{S_w}\boldsymbol{x} = \boldsymbol{x} + \frac{1}{\ell_w} \sum_{i=\ell_1+\ldots+\ell_{w-1}+1}^{\ell_1+\ldots+\ell_w} \frac{b_i - \langle \boldsymbol{a}^i, \boldsymbol{x} \rangle}{\|\boldsymbol{a}^i\|^2} \boldsymbol{a}^i, \tag{4}$$

where $\|\boldsymbol{a}\|$ denotes the norm of the vector $\boldsymbol{a}$ in $\mathbb{R}^J$, and $\mathbf{Q} : \Delta \to \Omega$ is defined by

$$(\mathbf{Q}\boldsymbol{x})_j = \max\{0, \boldsymbol{x}_j\}, \text{ for } 1 \leq j \leq J. \tag{5}$$

Note that $\mathbf{R}_S : \Delta \to \Omega$. This specific algorithm $\mathbf{R}$ is a typical example of the so-called block-iterative methods mentioned above. Except for the presence of $\mathbf{Q}$ in Eq. (3), which enforces non-negativity of the components, it is identical to an algorithm used and illustrated in Ref. 31. With the $\mathbf{Q}$ absent from the definition of the algorithm, Ω has to be the whole of $\mathbb{R}^J$; the practical consequence of the presence versus the absence of $\mathbf{Q}$ in the tomographic application is illustrated in Subsection III.D. We also note that special cases of the presented algorithm include the classical reconstruction methods such as algebraic reconstruction technique (ART) (if $\ell_w = 1$, for $1 \leq w \leq W$) and SIRT (if $W = 1$); see, for example, Chaps. 11 and 12 of Ref. 55.

For a problem structure $\langle \mathbb{T}, \mathcal{P}r \rangle$, a $T \in \mathbb{T}$, an $\varepsilon \in \mathbb{R}_+$, and a sequence $R = (\boldsymbol{x}^k)_{k=0}^{\infty}$ of points in Ω, we use $O(T, \varepsilon, R)$ to denote the $\boldsymbol{x} \in \Omega$ that has the following properties: $\mathcal{P}r_T(\boldsymbol{x}) \leq \varepsilon$ and there is a non-negative integer K such that $\boldsymbol{x}^K = \boldsymbol{x}$ and, for all non-negative integers $k < K \mathcal{P}r_T(\boldsymbol{x}^k) > \varepsilon$. Clearly, if there is such an $\boldsymbol{x}$, then it is unique. If there is no such $\boldsymbol{x}$, then we say that $O(T, \varepsilon, R)$ is *undefined,* otherwise we say that it is *defined*. The intuition behind this definition is the following: if we think of R as the (infinite) sequence of points that is produced by an algorithm (intended for the problem T) without a termination criterion, then $O(T, \varepsilon, R)$ is the *output* produced by that algorithm when we add to it instructions that make it terminate as soon as it reaches a point that is ε-compatible with T.

II.C. Bounded perturbation resilience

The notion of a *bounded perturbations resilient* algorithm $\mathbf{P}$ for a problem structure $\langle \mathbb{T}, \mathcal{P}r \rangle$ has been defined in a mathematically precise manner.[29] However, that definition is not satisfactory from the point of view of applications in medical physics (or indeed in any area involving noisy data), because it is useful only for problems T for which there is a perfectly compatible solution (that is, an $\boldsymbol{x}$ such that $\mathcal{P}r_T(\boldsymbol{x}) = 0$). We therefore extend here that notion as follows. An algorithm $\mathbf{P}$ for a problem structure $\langle \mathbb{T}, \mathcal{P}r \rangle$ is said to be *strongly perturbation resilient* if, for all $T \in \mathbb{T}$,

(i) there exists an $\varepsilon \in \mathbb{R}_+$ such that $O(T, \varepsilon, ((\mathbf{P}_T)^k \boldsymbol{x})_{k=0}^{\infty})$ is defined for every $\boldsymbol{x} \in \Omega$;

(ii) for all $\varepsilon \in \mathbb{R}_+$ such that $O(T, \varepsilon, ((\mathbf{P}_T)^k \boldsymbol{x})_{k=0}^{\infty})$ is defined for every $\boldsymbol{x} \in \Omega$, we also have that $O(T, \varepsilon', R)$ is defined for every $\varepsilon' > \varepsilon$ and for every sequence $R = (\boldsymbol{x}^k)_{k=0}^{\infty}$ of points in Ω generated by

$$\boldsymbol{x}^{k+1} = \mathbf{P}_T(\boldsymbol{x}^k + \beta_k \boldsymbol{v}^k), \text{ for all } k \geq 0, \tag{6}$$

where $\beta_k \boldsymbol{v}^k$ are *bounded perturbations*, meaning that the sequence $(\beta_k)_{k=0}^{\infty}$ of non-negative real numbers is *summable* (that is, $\sum_{k=0}^{\infty} \beta_k < \infty$), the sequence $(\boldsymbol{v}^k)_{k=0}^{\infty}$ of vectors in $\mathbb{R}^J$ is bounded and, for all $k \geq 0$, $\boldsymbol{x}^k + \beta_k \boldsymbol{v}^k \in \Delta$.

In less formal terms, the second of these properties says that for a strongly perturbation resilient algorithm we have that, for every problem and any non-negative real number ε, if it is the case that for all initial points from Ω the infinite sequence produced by the algorithm contains an ε-compatible point, then it will also be the case that all perturbed sequences satisfying Eq. (6) contain an ε'-compatible point, for any $\varepsilon' > \varepsilon$.

Having defined the notion of a strongly perturbation resilient algorithm, we next show that this notion is of relevance to problems in medical physics. We illustrate the use of this in tomography in Sec. III. We first need to introduce some mathematical concepts.

Given an algorithm $\mathbf{P}$ for a problem structure $\langle \mathbb{T}, \mathcal{P}r \rangle$ and a $T \in \mathbb{T}$, we say that $\mathbf{P}$ is *convergent for* T if, for every $\boldsymbol{x} \in \Omega$, there exists a unique $\boldsymbol{y}(\boldsymbol{x}) \in \Omega$ such that, $\lim_{k\to\infty} (\mathbf{P}_T)^k \boldsymbol{x} = \boldsymbol{y}(\boldsymbol{x})$, meaning that for every positive real number δ, there exists a non-negative integer K, such that $\|(\mathbf{P}_T)^k \boldsymbol{x} - \boldsymbol{y}(\boldsymbol{x})\| \leq \delta$, for all non-negative integers $k \geq K$. If, in addition, there exists a $\gamma \in \mathbb{R}_+$ such that $\mathcal{P}r_T(\boldsymbol{y}(\boldsymbol{x})) \leq \gamma$, for every $\boldsymbol{x} \in \Omega$, then we say that $\mathbf{P}$ is *boundedly convergent for* T.

A function $f : \Omega \to \mathbb{R}$ is *uniformly continuous* if, for every $\varepsilon > 0$ there exists a $\delta > 0$, such that, for all $\boldsymbol{x}, \boldsymbol{y} \in \Omega$, $|f(\boldsymbol{x}) - f(\boldsymbol{y})| \leq \varepsilon$ provided that $\|\boldsymbol{x} - \boldsymbol{y}\| \leq \delta$. An example of a uniformly continuous function is Res_S of Eq. (2), for any $S \in \mathbb{S}$. This can be proved by observing that the right-hand side of Eq. (2) can be rewritten in vector/matrix form as $\|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|$ and then selecting, for any given $\varepsilon > 0$, δ to be $\varepsilon / \|\boldsymbol{A}\|$, where $\|\boldsymbol{A}\|$ denotes the matrix norm of $\boldsymbol{A}$.

An operator $\mathbf{O}$: $\Delta \to \Omega$, is *nonexpansive* if $\|\mathbf{O}\boldsymbol{x} - \mathbf{O}\boldsymbol{y}\| \leq \|\boldsymbol{x} - \boldsymbol{y}\|$, for all $\boldsymbol{x}, \boldsymbol{y} \in \Delta$. An example of a nonexpansive operator is the $\mathbf{R}_S$ of Eq. (3). The proof of this is also simple. It follows from discussions regarding similar claims in Ref. 27 that the $\mathbf{B}_{S_w} : \mathbb{R}^J \to \mathbb{R}^J$ of Eq. (4) is a nonexpansive operator, for $1 \leq w \leq W$, and that the operator $\mathbf{Q}$ of Eq. (5) is also nonexpansive. Obviously, a sequential application of nonexpansive operators results in a nonexpansive operator and thus $\mathbf{R}_S$ is nonexpansive.

Now we state an important new result that gives sufficient conditions for strong perturbation resilience: If $\mathbf{P}$ is an algorithm for a problem structure $\langle \mathbb{T}, \mathcal{P}r \rangle$ such that, for all $T \in \mathbb{T}$, $\mathbf{P}$ is boundedly convergent for T, $\mathcal{P}r_T : \Omega \to \mathbb{R}$ is uniformly continuous, and $\mathbf{P}_T : \Delta \to \Omega$ is nonexpansive, then $\mathbf{P}$ is strongly perturbation resilient. The importance of this result lies in the fact that the rather ordinary condition of uniform continuity for the proximity function and the reasonable conditions of bounded convergence and nonexpansiveness of the algorithmic operators guarantee that we end up with a strongly perturbation resilient algorithm. The proof of this new result involves some mathematical technicalities and is therefore presented in the Appendix as Theorem 1.

II.D. Optimization criterion and nonascending vector

Now suppose, as is indeed the case for the constrained optimization problems discussed in Sec. I, that in addition to a problem structure $\langle \mathbb{T}, \mathcal{P}r \rangle$ we are also provided with an optimization criterion, which is specified by a function $\phi : \Delta \to \mathbb{R}$, with the convention that a point in Δ for which the value of ϕ is smaller is considered *superior* (from the point of view of our application) to a point in Δ for which the value of ϕ is larger. In the tomography context, any of the functions of $\boldsymbol{x}$ that are listed as a "secondary optimization criterion" (an alternative name is a "regularizer") in Sec. 6.4 of Ref. 55 is an acceptable choice for the optimization criterion ϕ. These include weighted norms, the negative of Shannon's entropy and total variation. It is the last of these that we discuss in detail in the illustrative example below. The essential idea of the *superiorization methodology* presented in this paper is to make use of the perturbations of Eq. (6) to transform a strongly perturbation resilient algorithm that seeks a constraints-compatible solution into one whose outputs are equally good from the point of view of constraints-compatibility, but are superior according to the optimization criterion. We do this by producing from the algorithm another one, called its *superiorized* version, by making sure not only that the $\beta_k \boldsymbol{v}^k$ are bounded perturbations, but also that $\phi(\boldsymbol{x}^k + \beta_k \boldsymbol{v}^k) \leq \phi(\boldsymbol{x}^k)$, for all $k \geq 0$.

In order to ensure this we introduce a new concept (closely related to the concept of a "descent direction" that is widely used in optimization). Given a function $\phi : \Delta \to \mathbb{R}$ and a point $\boldsymbol{x} \in \Delta$, we say that a vector $\boldsymbol{d} \in \mathbb{R}^J$ is *nonascending* for ϕ at $\boldsymbol{x}$ if $\|\boldsymbol{d}\| \leq 1$ and

$$\begin{aligned} &\text{there is a } \delta > 0 \text{ such that for all } \lambda \in [0, \delta], \\ &\quad (\boldsymbol{x} + \lambda \boldsymbol{d}) \in \Delta \text{ and } \phi(\boldsymbol{x} + \lambda \boldsymbol{d}) \leq \phi(\boldsymbol{x}). \end{aligned} \tag{7}$$

Note that irrespective of the choices of ϕ and $\boldsymbol{x}$, there is always at least one nonascending vector $\boldsymbol{d}$ for ϕ at $\boldsymbol{x}$, namely, the zero-vector, all of whose components are zero. This is a useful fact for proving results concerning the guaranteed behavior of our proposed procedures. However, in order to steer our algorithms towards a point at which the value of ϕ is small, we need to find a $\boldsymbol{d}$ such that $\phi(\boldsymbol{x} + \lambda \boldsymbol{d}) < \phi(\boldsymbol{x})$ rather than just $\phi(\boldsymbol{x} + \lambda \boldsymbol{d}) \leq \phi(\boldsymbol{x})$ as in Eq. (7). In some earlier papers on superiorization[27–31] it was assumed that $\Delta = \mathbb{R}^J$ and that ϕ is a convex function. This implied that, for any point $\boldsymbol{x} \in \Delta$, ϕ had a subgradient $\boldsymbol{g} \in \mathbb{R}^J$ at the point $\boldsymbol{x}$. It was suggested that if there is such a $\boldsymbol{g}$ with a positive norm, then $\boldsymbol{d}$ should be chosen to be $-\boldsymbol{g}/\|\boldsymbol{g}\|$, otherwise $\boldsymbol{d}$ should be chosen to be the zero vector. However, there are approaches (not involving subgradients) to selecting an appropriate $\boldsymbol{d}$; an example can be found in Ref. 32 in which $\boldsymbol{d}$ is found without using subgradients for the case when ϕ is the ℓ_1-norm of the Haar transform.

The method we used for selecting a nonascending vector in the experiments reported in this paper is specified at the end of Subsection III.A.

II.E. Superiorized version of an algorithm

We now make precise the ingredients needed for transforming an algorithm into its superiorized version. Let Ω and Δ be the underlying sets for a problem structure $\langle \mathbb{T}, \mathcal{P}r \rangle$ ($\Omega \subseteq \Delta \subseteq \mathbb{R}^J$, as discussed at the beginning of Subsection II.B), $\mathbf{P}$ be an algorithm for $\langle \mathbb{T}, \mathcal{P}r \rangle$ and $\phi : \Delta \to \mathbb{R}$. The following description of the Superiorized Version of Algorithm $\mathbf{P}$ produces, for any problem $T \in \mathbb{T}$, a sequence $R_T = (\boldsymbol{x}^k)_{k=0}^{\infty}$ of points in Ω for which, for all $k \geq 0$, Eq. (6) is satisfied. We show this to be true, for any algorithm $\mathbf{P}$, after the description of the Superiorized Version of Algorithm $\mathbf{P}$. Furthermore, since the sequence R_T is steered by Superiorized Version of Algorithm $\mathbf{P}$ towards a reduced value of ϕ, there is an intuitive expectation that the output of the superiorized version is likely to be superior (from the point of view of the optimization criterion ϕ) to the output of the original unperturbed algorithm. This last statement is not precise and so it cannot be proved in a mathematical sense for an arbitrary algorithm $\mathbf{P}$; however, that should not stop us from applying the easy procedure given below for automatically producing the superiorized version of $\mathbf{P}$ and experimentally checking whether it indeed provides us with outputs superior to those of the original algorithm. The well-demonstrated nature of heuristic optimization approaches is that they often work in practice even when their performance cannot be guaranteed to be optimal.[33–35]

Nevertheless, we can push our theory further than the hope expressed in the last paragraph, by considering superiorized versions of algorithms that satisfy some condition. In this paper, the condition that we discuss is strong perturbation resilience. We show below that if $\mathbf{P}$ is strongly perturbation resilient, then, for any problem $T \in \mathbb{T}$, a sequence R_T produced by its superiorized version has the following desirable property: For all $\varepsilon \in \mathbb{R}_+$, if $O(T, \varepsilon, ((\mathbf{P}_T)^k \boldsymbol{x})_{k=0}^{\infty})$ is defined for every $\boldsymbol{x} \in \Omega$, then $O(T, \varepsilon', R_T)$ is also defined for every $\varepsilon' > \varepsilon$; in other words, the Superiorized Version of Algorithm $\mathbf{P}$ provides an ε'-compatible output. As stated above, the advantage of the superiorized version is that its output is likely to be superior to the output of the original unperturbed algorithm. We point out that strong perturbation resilience is a sufficient, but not necessary, condition for guaranteeing such desirable behavior of the superiorized version, finding additional sufficient conditions and proving that algorithms that we wish to superiorize satisfy such conditions is part of our ongoing research.

The superiorized version assumes that we have available a summable sequence $(\gamma_\ell)_{\ell=0}^{\infty}$ of positive real numbers (for example, $\gamma_\ell = a^\ell$, where $0 < a < 1$) and it generates, simultaneously with the sequence $(\boldsymbol{x}^k)_{k=0}^{\infty}$, sequences $(\boldsymbol{v}^k)_{k=0}^{\infty}$, and $(\beta_k)_{k=0}^{\infty}$. The latter is generated as a subsequence of $(\gamma_\ell)_{\ell=0}^{\infty}$, resulting in a summable sequence $(\beta_k)_{k=0}^{\infty}$. The algorithm further depends on a specified initial point $\bar{\boldsymbol{x}} \in \Omega$ and on a positive integer N. It makes use of a logical variable called *loop*.

Superiorized Version of Algorithm P

(i) **set** $k = 0$
(ii) **set** $\boldsymbol{x}^k = \bar{\boldsymbol{x}}$
(iii) **set** $\ell = -1$
(iv) **repeat**
(v) **set** $n = 0$
(vi) **set** $\boldsymbol{x}^{k,n} = \boldsymbol{x}^k$
(vii) **while** $n < N$
(viii) **set** $\boldsymbol{v}^{k,n}$ to be a nonascending vector for ϕ at $\boldsymbol{x}^{k,n}$
(ix) **set** *loop* = *true*
(x) **while** *loop*
(xi) **set** $\ell = \ell + 1$
(xii) **set** $\beta_{k,n} = \gamma_\ell$
(xiii) **set** $z = \boldsymbol{x}^{k,n} + \beta_{k,n} \boldsymbol{v}^{k,n}$
(xiv) **if** $z \in \Delta$ **and** $\phi(z) \leq \phi(\boldsymbol{x}^k)$, **then**
(xv) **set** $n = n + 1$
(xvi) **set** $\boldsymbol{x}^{k,n} = z$
(xvii) **set** *loop* = *false*
(xviii) **set** $\boldsymbol{x}^{k+1} = \mathbf{P}_T \boldsymbol{x}^{k,N}$
(xix) **set** $k = k + 1$.

Next we analyze the behavior of the Superiorized Version of Algorithm $\mathbf{P}$.

The iteration number k is set to 0 in (i) and $\boldsymbol{x}^k = \boldsymbol{x}^0$ is set to its initial value $\bar{\boldsymbol{x}}$ in (ii). The integer index ℓ for picking the next element from the sequence $(\gamma_\ell)_{\ell=0}^{\infty}$ is initialized to -1 by line (iii), it is repeatedly increased by line (xi). The lines (v)–(xix) that follow the **repeat** in (iv) perform a complete iterative step from $\boldsymbol{x}^k$ to $\boldsymbol{x}^{k+1}$, infinite repetitions of such steps provide the sequence $R_T = (\boldsymbol{x}^k)_{k=0}^{\infty}$. During one iterative step, there is one application of the operator $\mathbf{P}_T$, in line (xviii), but there are N steering steps aimed at reducing the value of ϕ; the latter are done by lines (v)–(xvii). These lines produce a sequence of points $\boldsymbol{x}^{k,n}$, where $0 \leq n \leq N$ with $\boldsymbol{x}^{k,0} = \boldsymbol{x}^k$, $\boldsymbol{x}^{k,n} \in \Delta$, and $\phi(\boldsymbol{x}^{k,n}) \leq \phi(\boldsymbol{x}^k)$.

We prove the truth of the last sentence by induction on the non-negative integers. For $n = 0$, we have by lines (v) and (vi) that $\boldsymbol{x}^{k,0} = \boldsymbol{x}^k$. But $\boldsymbol{x}^k \in \Omega$, since it is either $\bar{\boldsymbol{x}}$ that is assumed to be in Ω due to lines (i) and (ii) or it is in the range Ω of $\mathbf{P}_T$ due to lines (xviii) and (xix). Now we assume, for any $0 \leq n < N$, that $\boldsymbol{x}^{k,n} \in \Delta$ and $\phi(\boldsymbol{x}^{k,n}) \leq \phi(\boldsymbol{x}^k)$ and show that lines (viii)–(xvii) perform a computation that leads from $\boldsymbol{x}^{k,n}$ to an $\boldsymbol{x}^{k,n+1} \in \Delta$ that satisfies $\phi(\boldsymbol{x}^{k,n+1}) \leq \phi(\boldsymbol{x}^k)$. To see this, observe that line (viii) sets $\boldsymbol{v}^{k,n}$ to be a nonascending vector for ϕ at $\boldsymbol{x}^{k,n}$, which implies that Eq. (7) is satisfied with $\boldsymbol{x} = \boldsymbol{x}^{k,n}$ and $\boldsymbol{d} = \boldsymbol{v}^{k,n}$. Line (ix) sets *loop* to *true*, and it remains *true* while searching for the desired $\boldsymbol{x}^{k,n+1}$, by repeatedly executing the loop sequence that follows line (x). In this sequence, line (xi) increases ℓ by 1 and line (xii) sets $\beta_{k,n}$ to γ_ℓ. Thus for the vector z defined by line (xiii), $z \in \Delta$ and $\phi(z) \leq \phi(\boldsymbol{x}^{k,n})$, provided that $\beta_{k,n}$ is not greater than the δ in Eq. (7). Since $(\gamma_\ell)_{\ell=0}^{\infty}$ is a summable sequence of positive real numbers, there must be a positive integer L such that $\gamma_\ell \leq \delta$, for all $\ell \geq L$. This implies that if we applied lines (xi)–(xiii) often enough, we would reach a vector z that satisfies $z \in \Delta$ and $\phi(z) \leq \phi(\boldsymbol{x}^{k,n})$. If the condition in line (xiv) is not satisfied when the process gets to it, then lines

(xi)–(xiii) are again executed and eventually we get a vector $\boldsymbol{z}$ for which the condition in line (xiv) is satisfied due to the induction hypothesis that $\phi(\boldsymbol{x}^{k,n}) \leq \phi(\boldsymbol{x}^k)$. By lines (xv) and (xvi) we see that at that time $\boldsymbol{x}^{k,n+1}$ is set to $\boldsymbol{z}$ and so we obtain that $\boldsymbol{x}^{k,n+1} \in \Delta$ and $\phi(\boldsymbol{x}^{k,n+1}) \leq \phi(\boldsymbol{x}^k)$, as desired. Line (xvii) sets *loop* to *false* and so control is returned to line (vii). When this happens for the Nth time, it will be the case that $n = N$ and, therefore, line (xviii) is used to produce $\boldsymbol{x}^{k+1} \in \Omega$ and the increasing of k by line (xix) allows us then to move on to the next iterative step. Infinite repetition of such steps produces the sequence $R_T = (\boldsymbol{x}^k)_{k=0}^{\infty}$ of points in Ω.

We now show that if $O(T, \varepsilon, ((\mathbf{P}_T)^k \boldsymbol{x})_{k=0}^{\infty})$ is defined for every $\boldsymbol{x} \in \Omega$, then, for any $\varepsilon' > \varepsilon$, the Superiorized Version of Algorithm $\mathbf{P}$ produces an ε'-compatible output. Since $\mathbf{P}$ is assumed to be strongly perturbation resilient, this desired result follows if we can show that there exists a summable sequence $(\beta_k)_{k=0}^{\infty}$ of non-negative real numbers and a bounded sequence $(\boldsymbol{v}^k)_{k=0}^{\infty}$ of vectors in $\mathbb{R}^J$ such that Eq. (6) is satisfied for all $k \geq 0$. In view of line (xviii), this is achieved if we can define the β_k and the $\boldsymbol{v}^k$ so that $\boldsymbol{x}^{k,N} = \boldsymbol{x}^k + \beta_k \boldsymbol{v}^k$. This is done by setting

$$\beta_k = \max\{\beta_{k,n} \,|\, 0 \leq n < N\}, \tag{8}$$

$$\boldsymbol{v}^k = \sum_{n=0}^{N-1} \frac{\beta_{k,n}}{\beta_k} \boldsymbol{v}^{k,n}. \tag{9}$$

That these assignments result in $\boldsymbol{x}^{k,N} = \boldsymbol{x}^k + \beta_k \boldsymbol{v}^k$ follows from lines (v)–(xvii). From line (xii) follows that $(\beta_k)_{k=0}^{\infty}$ is a subsequence of $(\gamma_\ell)_{\ell=0}^{\infty}$ and, hence, it is a summable sequence of non-negative real numbers. Since each $\|\boldsymbol{v}^{k,n}\| \leq 1$ by the definition of a nonascending vector, it follows from Eqs. (8) and (9) that $\|\boldsymbol{v}^k\| \leq N$ and so $(\boldsymbol{v}^k)_{k=0}^{\infty}$ is bounded. Part of the condition expressed in Eq. (6) is that, for all $k \geq 0$, $\boldsymbol{x}^k + \beta_k \boldsymbol{v}^k \in \Delta$. This follows from the fact that $\boldsymbol{x}^{k,N} = \boldsymbol{x}^k + \beta_k \boldsymbol{v}^k$ is assigned its value by line (xvi), but only if the condition expressed in line (xiv) is satisfied.

In conclusion, we have shown that the superiorized version of a strongly perturbation resilient algorithm produces outputs that are essentially as constraints-compatible as those produced by the original version of the algorithm. However, due to the repeated steering of the process by lines (vii)–(xvii) towards reducing the value of the optimization criterion ϕ, we can expect that the output of the superiorized version will be superior (from the point of view of ϕ) to the output of the original algorithm.

II.F. Information on performance comparison with MAP methods

Using our notation, the constrained minimization formulation that we are considering is as follows: Given an $\varepsilon \in \mathbb{R}_+$,

$$\text{minimize } \phi(\boldsymbol{x}), \text{ subject to } \mathcal{P}r_T(\boldsymbol{x}) \leq \varepsilon. \tag{10}$$

The aim of superiorization is not identical with the aim of constrained minimization in Eq. (10). One difference is that ε is not "given" in the superiorization context. The superiorization of an algorithm produces a sequence and, for any ε, the associated output of the algorithm is considered to be the first $\boldsymbol{x}$ in the sequence for which $\mathcal{P}r_T(\boldsymbol{x}) \leq \varepsilon$. The other difference is that we do not claim that this output is a minimizer of ϕ among all points that satisfy the constraint, but hope only that it is usually an $\boldsymbol{x}$ for which $\phi(\boldsymbol{x})$ is at the small end of its range of values over the set of constraint-satisfying points. This latter difference is generally shared by comparisons of a heuristic approach with an exact approach to solving a constrained minimization problem.

The MAP (or regularized) formulation of a physical problem that leads to the constrained minimization problem (10) is the unconstrained minimization problem of the form: Given a $\beta \in \mathbb{R}_+$,

$$\text{minimize } [\phi(\boldsymbol{x}) + \beta \mathcal{P}r_T(\boldsymbol{x})]. \tag{11}$$

Formulations of both kinds [i.e., the ones of Eqs. (10) and (11)] are widely used for solving medical physics problems and the question "Which of these two formulations leads to faster or better solutions of the underlying physical problem?" is open. Examples of both formulations with various choices for $\mathcal{P}r_T$ and ϕ are listed in the beginning parts of the paper of Goldstein and Osher.[47]

We now return to the question raised near the end of Sec. I: Will superiorization produce superior results to those produced by contemporary MAP methods or is it faster than the better of such methods? As yet, there is very little information available regarding this general question; in fact, we are aware of only one published study.[45] That study compared a superiorization algorithm with the algorithm of Goldstein and Osher that they refer to as TwIST (Ref. 46) with split Bregman[47] as the substep, which is indeed a contemporary method that uses the MAP formulation. (For example, see the discussion of the split Bregman method in Ref. 56.) The problem S to which the two algorithms were applied was one from the tomographic problem set $\mathbb{S}$ defined in Eq. (1). Res_S as defined in Eq. (2) was used as the proximity function and total variation, TV as defined below in Eq. (12), was the choice for ϕ. It is reported in Ref. 45 that for the outputs of the two algorithms that were being compared, the values of Res_S and TV were very similar, but the superiorization algorithm produced its output four times faster than the MAP method.

III. AN ILLUSTRATIVE EXAMPLE

III.A. Application to tomography

We use *tomography* to refer to the process of reconstructing a function over a Euclidean space from estimated values of its integrals along lines (that are usually, but not necessarily, straight). The particular reconstruction processes to which our discussion applies are the *series expansion methods*, see Sec. 6.3 of Ref. 55, in which it is assumed that the function to be reconstructed can be approximated by a linear combination of a finite number (say J) of basis functions and the

reconstruction task becomes one of estimating the coefficients of the basis functions in the expansion. Sometimes, prior knowledge about the nature of the function to be reconstructed allows us to confine the sought-after vector $\boldsymbol{x}$ of coefficients to a subset Ω of $\mathbb{R}^J$ (such as the non-negative orthant $\mathbb{R}^J_+$). We use i to index the lines along which we integrate, $\boldsymbol{a}^i \in \mathbb{R}^J$ to denote the vector whose jth component is the integral of the jth basis function along the ith line, and b_i to denote the measured integral of the function to be reconstructed along the ith line. Under these circumstances the constraints come from the desire that, for each of the lines, $\langle \boldsymbol{a}^i, \boldsymbol{x} \rangle$ should be close (in some sense) to b_i.

To make this concrete, consider Eq. (1). Such a description of the constraints arises in tomography by grouping the lines of integration into W blocks, with ℓ_w lines in the wth block. Such groupings often (but not always) are done according to some geometrical condition on the lines (for example, in case of straight lines, we may decide that all the lines that are parallel to each other form one block). In this framework, the proximity function *Res* defined by Eq. (2) provides a reasonable measure of the incompatibility of a vector $\boldsymbol{x}$ with the constraints. The algorithm **R** described by Eqs. (3)–(5) is applicable to this concrete formulation.

There are many optimization criteria that have been used in tomography, see Sec. 6.4 of Ref. 55, here we discuss the one called TV, whose use has been popular in medical physics recently, see as examples Refs. 20, 22, 23, and 41–44. The definition of TV that we use here requires a certain way of selecting the basis functions. It is assumed that the function to be reconstructed is defined in the plane $\mathbb{R}^2$ and is zero-valued outside a square-shaped region in the plane. This region is subdivided into J smaller equal-sized squares (*pixels*) and the J basis functions are defined by having value one in exactly one pixel and value zero everywhere else. We index the pixels by j and we let C denote the set of all indices of pixels that are not in the rightmost column or the bottom row of the pixel array. For any pixel with index j in C, let $r(j)$ and $b(j)$ be the index of the pixel to its right and below it, respectively. We define $TV : \mathbb{R}^J \to \mathbb{R}$ by

$$TV(\boldsymbol{x}) = \sum_{j \in C} \sqrt{(x_j - x_{r(j)})^2 + (x_j - x_{b(j)})^2}. \tag{12}$$

The method we adopted to generate a nonascending vector for the TV function at an $\boldsymbol{x} \in \mathbb{R}^J$ is based on Theorem 2 of the Appendix. It is applicable since $TV : \mathbb{R}^J \to \mathbb{R}$ is a convex function; see, for example, the end of the Proof of Proposition 1 of Ref. 41. Now consider an integer j' such that $1 \le j' \le J$. Looking at the sum in Eq. (12), we see that $x_{j'}$ appears in at most three terms, in which j' must be either j, or $r(j)$, or $b(j)$ for some $j \in C$. By taking the formal partial derivatives of these three terms, we see that $\frac{\partial TV}{\partial x_{j'}}(\boldsymbol{x})$ is well defined if the denominator in the formal derivative of each of the three terms is not zero for $\boldsymbol{x}$. In view of this, we define the $\boldsymbol{g}$ in Theorem 2 as follows. If the denominator in any of the three formal partial derivatives with respect to $x_{j'}$ has an absolute value less than a very small positive number (we used 10^{-20}), then we set $g_{j'}$ to zero, otherwise we set it to $\frac{\partial TV}{\partial x_{j'}}(\boldsymbol{x})$. Clearly, the resulting $\boldsymbol{g} \in \mathbb{R}^J$ satisfies the condition in Theorem 2 and hence provides a $\boldsymbol{d}$ that is a nonascending vector for TV at $\boldsymbol{x}$.

Previously reported reconstructions using TV-superiorization selected the $\boldsymbol{d}$ using subgradients as discussed in the paragraph following Eq. (7); such a $\boldsymbol{d}$ is not guaranteed to be a nonascending vector for the TV function. What we are proposing here is not only mathematically rigorous (in the sense that it is guaranteed to produce a nonascending vector for the TV function), but it can also lead to a better reconstructions, as illustrated in Subsection III.D.

III.B. The data generation for the experiments

The datasets used in the experiments reported in this paper were generated in such a way that they share the noise-characteristics of CT scanners when used for scanning the human head and brain; as discussed, for example, in Chap. 5 of Ref. 55. They were generated using the software SNARK09.[57]

The head phantom that was used for data generation is based on an actual cross section of the human head. It is described as a collection of geometrical objects (such as ellipses, triangles, and segments of circles) whose combination accurately resembles the anatomical features of the actual head cross section. In addition, the basic phantom contains a large tumor. The actual phantom used was obtained by a random variation of the basic phantom, by incorporating into it local inhomogeneities and small low-contrast tumors at random locations. This phantom is represented by the image in Fig. 1(a). That image comprises 485×485 pixels each of size 0.376 mm by 0.376 mm. The values assigned to the pixels are obtained by an 11×11 subsampling of the pixels and averaging the values assigned to the subsamples by the geometrical objects that are used to describe the anatomical features and the tumors. Those values are approximate linear attenuation coefficients per cm at 60 keV (0.416 for bone, 0.210 for brain, 0.207 for cerebrospinal fluid). The contrast of the small tumors with their background is 0.003 cm^{-1}. In order to clearly see the low-contrast details in the interior of the skull, we use zero (black) to represent the value 0.204 (or anything less) and 255 (white) to represent 0.21675 or anything more). The same is true for all the images in the rest of this paper.

For the selected head phantom we generated *parallel projection data*, in which one *view* comprises estimates of integrals through the phantom for a set of 693 equally spaced parallel lines with a spacing of 0.0376 cm between them. (We chose to simulate parallel rather than divergent projection data, since the reconstruction by the method of Ref. 42 with which we wish to compare the superiorization approach was performed for us by the authors of Ref. 42 on parallel data. Even though contemporary CT scanners use divergent projection data, results obtained by the use of parallel projection data are relevant to them, since it is known that the quality of reconstructions from these two modes of data collection are very similar as long as the data generations use similar frequencies of sampling of lines and similar noise characteristics

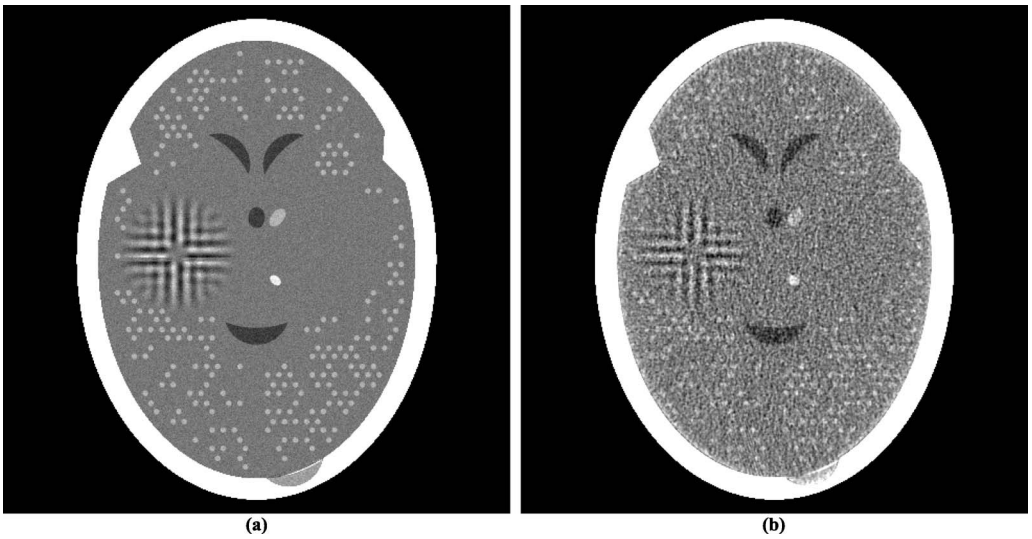


FIG. 1. (a) A head phantom. (b) Reconstruction of the head phantom from realistically simulated projection data for 360 views using ART with blob basis functions.

in the estimated integrals for those lines; see, for example, the reconstructions from divergent and parallel projection data in Fig. 5.15 of Ref. 55.) In calculating these estimates. we take into consideration the effects of photon statistics, detector width, and scatter. Details of how we do this exactly can be found in Secs. 5.5 and 5.9 of Ref. 55. Briefly, quantum noise is calculated based on the assumption that approximately 2 000 000 photons enter the head along each ray, detector width is simulated by using 11 subrays along each of which the attenuation is calculated independently and then combined at the detector, and 5% of the photons get counted not by the detector for the ray in question but detectors for the neighboring rays. For the experiments in this paper, we did not simulate the polyenergetic nature of the x-ray source. To indicate what can be achieved in clinical CT, we show in Fig. 1(b) a reconstruction that was made from data comprising of 360 such views with the reconstruction algorithm known as ART with blob basis functions; see Chap. 11 of Ref. 55.

III.C. Superiorization reconstruction from a few views

The main reason in the literature for advocating the use of TV as the optimization criterion is that by doing so one can achieve efficacious reconstructions even from sparsely sampled data. In our own work[31] with realistically simulated CT data, we found that this is not always the case and this will be demonstrated again by the experiments reported in the current paper.

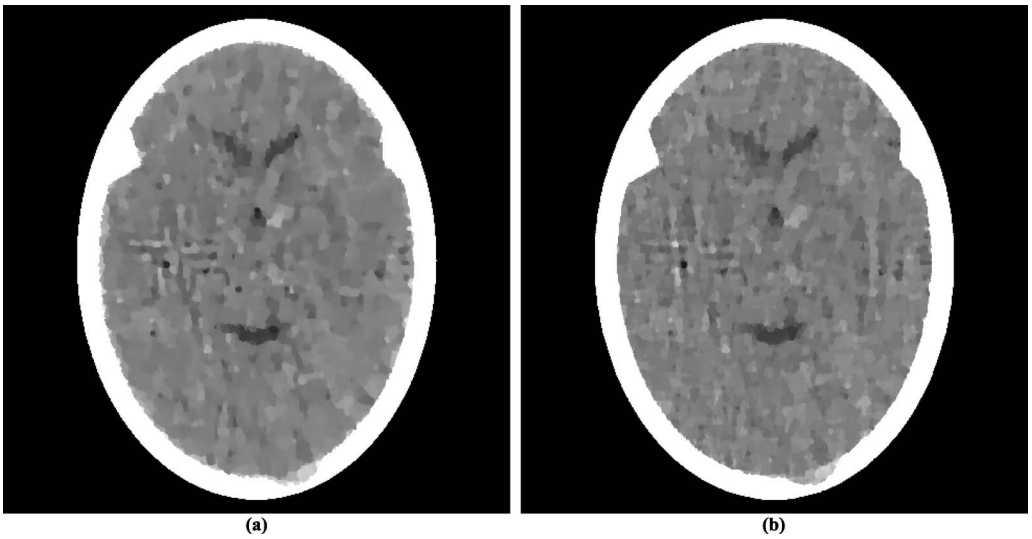


FIG. 2. Reconstructions using TV as the optimization criterion from realistically simulated projection data for 60 views using (a) ASD-POCS and (b) superiorization. As compared to Fig. 1(b), these reconstructions fail in two ways: they do not show some of the fine details in the phantom and they present some artifactual variations. The former of these is a consequence of reconstructing from a much smaller dataset than used for Fig. 1(b). The latter is due to using a very narrow window (13.5 HU) in these displays. Were we to use a wider display window (e.g., from –429 HU to 429 HU) for the reconstructions in this figure and in Fig. 1(b), the visual appearance of the resulting images would be nearly indistinguishable.

There have appeared in the literature some approaches to TV minimization that seem to indicate a more efficacious performance for CT than the one reported in Ref. 31. One of these is the adaptive steepest descent projections onto convex sets (ASD-POCS) algorithm, which is described in detail in the much-cited paper of Sidky and Pan[42] and whose use has been since reported in a number of subsequent publications, for example, in Refs. 23 and 43. We note that ASD-POCS was designed with the aim of producing an exact minimization algorithm, in contrast to our heuristic superiorization approach. Translating Eqs. (6)–(8) of Ref. 42 into our terminology, the aim of ASD-POCS is the following: Given an $\varepsilon \in \mathbb{R}_+$, find an ε-compatible $\boldsymbol{x} \in \Omega = \mathbb{R}_+^J$ for which $TV(\boldsymbol{x})$ is minimal. [Note that this aim is a special case of the constrained optimization formulation presented in Eq. (10).] In order to test ASD-POCS, we generated realistic projection data as described in Subsection III.B but for only 60 views at 3° increments with the spacing between the lines for which integrals are estimated set at 0.752 mm. Thus the number of rays (and hence the number photons put into the head) in this dataset is a 12th of what it is in the dataset used to produce the reconstruction in Fig. 1(b). A reconstruction from these data was produced for us using ASD-POCS by the authors of Ref. 42 (this ensured that it does not suffer due to our misinterpretation of the algorithm or from our inappropriate choices of the free parameters), it is shown in Fig. 2(a).

Since the image quality of Fig. 2(a) is not anywhere near to that of Fig. 1(b), we present here a brief discussion as to why we are showing such images. Many publications in the recent medical imaging literature have claimed that medically efficacious reconstructions can be obtained by the use of TV-minimization from data as sparse as what was used to produce Fig. 2(a). (In fact, ASD-POCS was motivated and used with such an aim in mind.[23,42,43]) Such publications usually show reconstructions from sparse data as evidence for the validity of their claims. They can do this because in their presented illustrations the features that are observable in the reconstructions are usually much larger and/or of much higher contrast against their backgrounds than the small "tumors" in Fig. 1(a), which are perfectly visible in the reconstruction in Fig. 1(b), but are not detectable in the reconstruction from sparse data in Fig. 2(a). The reason why that reconstruction appears to be unacceptably bad is that the display window (from 0.204 cm^{-1} linear attenuation coefficient to 0.21675 cm^{-1} linear attenuation coefficient) is very narrow; it was selected to enhance the visibility of the small low-contrast tumors. The width of this window corresponds to about 13.5 Hounsfield units (HU). As compared to this, in their evaluation of sparse-view reconstruction from flat-panel-detector cone-beam CT, Bian *et al.*[43] use what they call a "soft-tissue grayscale window" (also a "narrow window") from –429 HU to 429 HU to display head phantom reconstructions. Using such a window for our reconstructions shown Figs. 2(a) and 1(b) would result in images that are nearly indistinguishable from each other. Thus reporting the images using such a display window is consistent with the claim that a *TV*-minimizing reconstruction from a few views is similar in quality to a more traditional reconstruction from many views. However, our much narrower display window reveals that this is not really so. We therefore continue using our much narrower window in what follows, since it clearly reveals the nature of the reconstructions being compared, warts and all.

While this ASD-POCS reconstruction is not as good as it should be for diagnostic CT of the brain (due to the sparsity of the data), it is visually better than the reconstruction using superiorization from similar data as reported in Ref. 31. We discuss the reasons for this in Subsection III.D. Here, we concentrate on examining whether one can achieve a reconstruction using superiorization that is as good as that produced by ASD-POCS from the same data.

For this we first need to examine the numerical properties of the ASD-POCS reconstruction. This reconstruction uses 485×485 pixels each of size 0.376 mm by 0.376 mm. This implies that $J = 235{,}225$ and it also determines the components of the vectors $\boldsymbol{a}^i \in \mathbb{R}^J$ in the precise specification of the problem S. The Res_S, as defined by Eq. (2), of the ASD-POCS reconstruction is 0.33 and the TV, as defined by Eq. (12), is 835.

We applied to the same problem S a superiorized version of the algorithm **R** defined by Eq. (3). To complete the specification of **R**, we point out that for the ordering of views we chose the "efficient" one that was introduced in Ref. 58 and is also discussed on p. 209 of Ref. 55. The choices we made for the superiorization are the following: $\gamma_\ell = 0.99995^\ell$, $\bar{\boldsymbol{x}}$ is the zero vector, and $N = 20$. The nonascending vector was computed by the method described in the paragraph below [Eq. (12)]. Denoting by R_S the infinite sequence of points in Ω that is produced by the superiorized version of the algorithm **R** when applied to the problem S, we chose as our reconstruction $\boldsymbol{x}^* = O(S, 0.33, R_S)$. For such a reconstruction we have, by the definition of O, that $Res_S(\boldsymbol{x}^*) \leq 0.33$; in other words, the output of the superiorization algorithm is at least as constraints-compatible with S as the output of ASD-POCS. From the point of view of TV-minimization, our $\boldsymbol{x}^*$ is slightly better: $TV(\boldsymbol{x}^*) = 826$.

The superiorization reconstruction is displayed in Fig. 2(b). Visually, it is similar to the reconstruction produced by ASD-POCS. From the optimization point of view it achieves the desired aim better than ASD-POCS does, since it results in smaller values for both Res_S and for TV, even though only slightly.

That the two reconstructions in Fig. 2 are very similar is not surprising because a comparison of the pseudocodes reveals that the ASD-POCS algorithm in Ref. 42 is essentially a special case of the Superiorized Version of Algorithm **P**, even though it has been derived from rather different principles. To obtain the ASD-POCS algorithm from our methodology described here, we would have to choose ART (see Chap. 11 of Ref. 55) as the algorithm that we are superiorizing. Such a superiorization of ART was reported in the earliest paper on superiorization.[27] For the illustration in our current paper, we decided to superiorize the block-iterative algorithm **R** defined by Eq. (3). This illustrates the generality of the superiorization approach: it is applicable not only to a large class of constrained optimization problems, but also enables the use of any of a large class of iterative algorithms designed to

produce a constraints-compatible solutions. A recent publication aimed at producing an exact TV-minimizing algorithm based on the block-iterative approach is Ref. 44.

III.D. Effects of variations in the reconstruction approach

The reconstruction in Fig. 2(a) produced by ASD-POCS definitely "looks better" than a reconstruction in Ref. 31, which was obtained using superiorization from similar data. Since, as discussed in the last paragraph of Subsection III.C, the ASD-POCS algorithm in Ref. 42 can be obtained as a special case of superiorization, it must be that some of the choices made in the details of the implementations are responsible for the visual differences. An analysis of the implementational details adopted by the two approaches revealed several differences. After removing these differences, the superiorization approach produced the image in Fig. 2(b), which is very similar to the reconstruction produced by ASD-POCS. We now list the implementational choices that were made for superiorization to make its performance match that of the reported implementation of ASD-POCS.

One implementational difference is in the stopping-rule of the iterative algorithm; that is, the choice of ε in determining the output $O(S, \varepsilon, R_S)$. Since the data are noisy, the phantom itself does not match the data exactly. In previously reported implementations of superiorization it was assumed that the iterative process should terminate when an image is obtained that is approximately as constraints-compatible as the phantom; in the case of the phantom and the projections data on which we report here the value of Res_S for the phantom is approximately 0.91, which is larger than its value (0.33) for the reconstruction produced by ASD-POCS. The output $O(S, 0.91, R_S)$ is shown in Fig. 3(a). This is a wonderfully smooth reconstruction, its TV value is only 771. However, this smoothness comes at a price: we lose not only the ability to detect the large tumor, but we cannot even see anatomic features (such as the ventricular cavities) inside the brain. So it appears that, in order to see medically relevant features in the brain, *overfitting* (in the sense of producing a reconstruction from noisy data that is more constraints-compatible than the phantom) is desirable.

In the implementations that produced previously reported reconstructions by superiorization, the number N in the Superiorized Version of Algorithm **P** was always chosen to be 1. It is possible that this is the wrong choice, making only this change to what lead to the reconstruction in Fig. 2(b) results in the reconstruction shown in Fig. 3(b). That image appears similar to the image in Fig. 2(b), but it has a higher TV value, namely, 832, which is still very slightly lower than that of the ASD-POCS reconstruction. The choice $N = 20$ was based on the desire to maintain consistency with what has been practiced using ASD-POCS, see p. 4790 of Ref. 42. It appears that in the context of our paper the additional computing cost due to choosing N to be 20 rather than 1 is not really justified. (We note that if $\boldsymbol{d}$ is selected using subgradients as discussed in the paragraph following Eq. (7) and thus $\boldsymbol{d}$ is not guaranteed to be a nonascending vector for the TV function, then the choice of 20 rather than 1 for N results in a considerable improvement. However, an even greater improvement is achieved even with $N = 1$ by selecting $\boldsymbol{d}$ as recommended in this paper.)

Another important difference between the ASD-POCS implementation and the previous implementations of the superiorization approach is the size of the pixels in the reconstructions. For the ASD-POCS reconstruction this was selected to be 0.376 mm by 0.376 mm. In previously reported reconstructions by superiorization it was assumed that the edge of a pixel should be the same as the distance between the parallel lines along which the data are collected; that is, 0.752 mm for our problem S. This assumption proved to be false. TV-minimization takes care of undesirable artifacts that may otherwise arise due to the smaller pixels and this leads to a visual improvement. A superiorizing reconstruction with the larger pixels, using $\varepsilon = 0.33$ and $N = 20$, is shown in Fig. 3(c). (We note that the use of smaller pixels during iterative x-ray CT reconstructions was also suggested in Ref. 59. However, that approach is quite different from what is presented here: its final result uses larger pixels whose values are obtained by averaging assemblies of values provided by the iterative process to the smaller pixels. There is no such downsampling in our approach, our final result is presented using the smaller pixels. Its smoothness is due to reduction of TV by the superiorization approach rather than to averaging pixel values in a denser digitization.)

Combining the use of the larger pixels with $\varepsilon = 0.91$ and $N = 1$ results in the reconstruction shown in Fig. 3(d). This reconstruction, for which the superiorization options were selected according to what was done in Ref. 31, is visually inferior to those shown in our Fig. 2. The reconstructions displayed in Fig. 3 also illustrate another important point, namely, that even though the mathematical results discussed in this paper are valid for a large range of choices of the parameters in the superiorization algorithms, for medical efficacy of the reconstructions attention has to be paid to these choices since they can have a drastic effect on the quality of the reconstruction.

It has been mentioned in Subsection II.B that except for the presence of **Q** in Eq. (3), which enforces non-negativity of the components, **R** is identical to the algorithm used and illustrated in Ref. 31. It is known that CT reconstruction of the brain from many views does not suffer from ignoring the fact that the components of the $\boldsymbol{x}$, which represent linear attenuation coefficients, should be non-negative; as is illustrated in Fig. 1(b). This remains so when reconstructing from a few views using the method and data that we have been discussing: if we do everything in exactly the same way as was done to obtain the reconstruction with TV value 826 that is shown in our Fig. 2(b) but remove **Q** from Eq. (3), then we obtain a reconstruction in Fig. 4(a) whose TV value is 829.

Another variation that deserves discussion, because it has been suggested in the literature,[22] is one that does not come about by making choices for the general approach of the Superiorized Version of Algorithm **P** but rather by changing the nature of the approach. The variation in question is not applicable in general, but can be applied to the special case when the algorithm to be superiorized is the **R** defined by Eq. (3). It

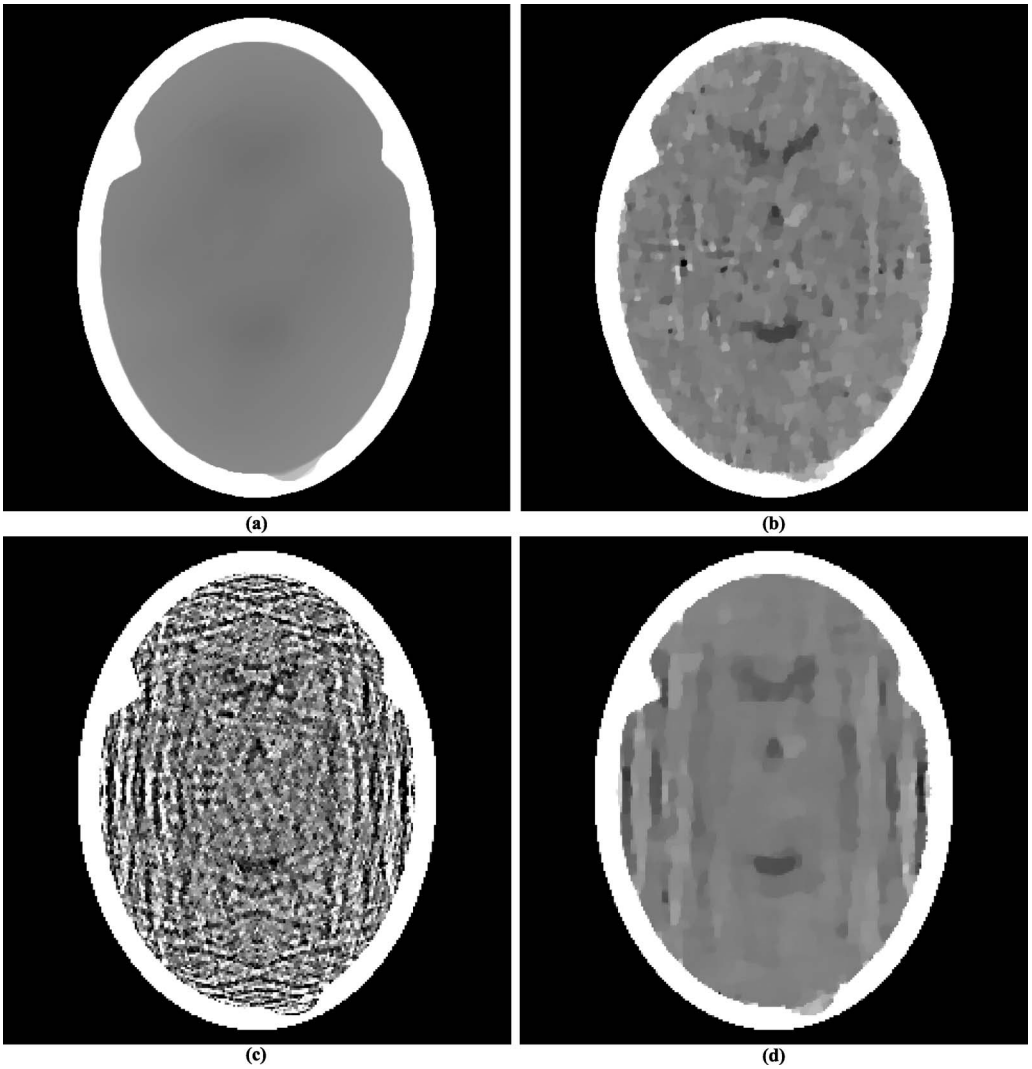


FIG. 3. Reconstructions produced by varying some of the parameters in the algorithm that produced Fig. 2(b). (a) Changing the termination criterion form $\varepsilon = 0.33$ to $\varepsilon = 0.91$. (b) Changing the value of N from 20 to 1. (c) Reconstructing with pixel size 0.752 mm by 0.752 mm instead of 0.376 mm by 0.376 mm. (d) Reconstructing with all the three changes of (a)–(c).

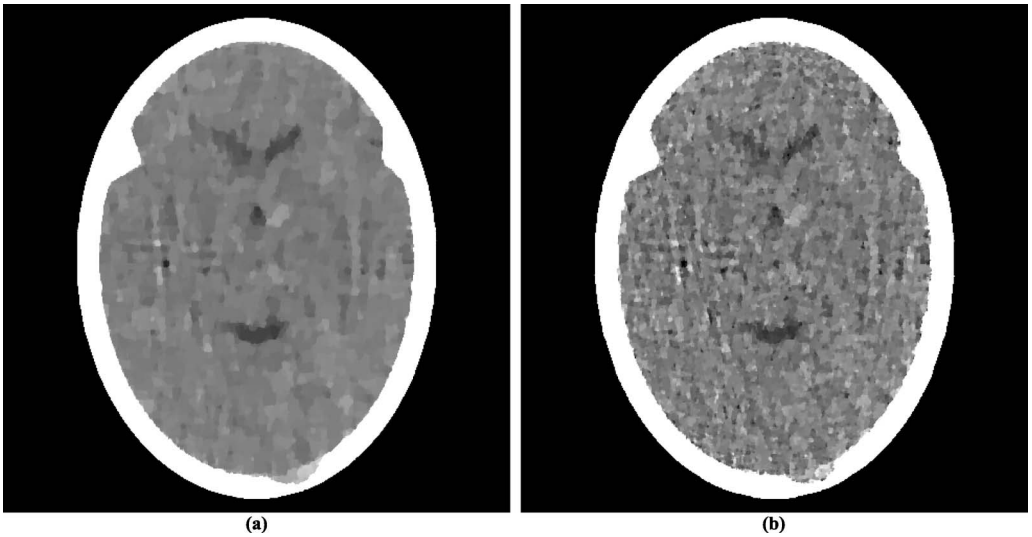


FIG. 4. Reconstructions by variations that do not fit into the framework within which the previously shown reconstructions were produced. (a) Not using non-negativity in the algorithm. (b) Interleaving perturbations with blocks.

was suggested as an improvement to the approach presented above with the choice $N = 1$. The idea was based on recognizing the block-iterative nature of the algorithmic operator $\mathbf{R}_S$ in Eq. (3) and intermingling the perturbation steps of lines (vii)–(xvii) of the Superiorized Version of Algorithm **R** with the projection steps $\mathbf{B}_{S_1}, \ldots, \mathbf{B}_{S_W}$ of Eq. (3). It was reported in Ref. 22 that doing this is advantageous to using the Superiorized Version of Algorithm **R**. However, when we applied the variation of the Superiorized Version of Algorithm **R** that is proposed in Ref. 22 to the problem S that we have been using in this section, we ended up with the reconstruction in Fig. 4(b) whose TV value is 920. This is not as good as what was obtained using the version of the algorithm that produced the reconstruction in Fig. 2(b). We conclude that the variation suggested by Ref. 22, which does not fit into the theory of our paper, does not have an advantage over what we are proposing here, at least for the problem S that we have been discussing in this section. We conjecture that the improvement reported in Ref. 22 is due to selecting $\boldsymbol{d}$ using subgradients as discussed in the paragraph following Eq. (7) and, as discussed earlier, such an improvement is not obtained if $\boldsymbol{d}$ is selected by the more appropriate method recommended in this paper.

IV. DISCUSSION AND CONCLUSIONS

Constrained optimization is an often-used tool in medical physics. The methodology of superiorization is a heuristic (as opposed to exact) approach to constrained optimization.

Although the idea of superiorization was introduced in 2007 and its practical use has been demonstrated in several publications since, this paper is the first to provide a solid mathematical foundation to superiorization as applied to the noisy problems of the real world. These foundations include a precise definition of constraints-compatibility, the concept of a strongly perturbation resilient algorithm, simple conditions that ensure that an algorithm is strongly perturbation resilient, the superiorized version of an algorithm and the showing that the superiorized version of a strongly perturbation resilient algorithm produces outputs that are essentially as constraints-compatible as those produced by the original version but are likely to have a smaller value of the chosen optimization criterion.

The approach is very general. For any iterative algorithm **P** and for any optimization criterion ϕ for which we know how to produce nonascending vectors, the pseudocode given in Subsection II.E automatically provides the version of **P** that is superiorized for ϕ.

We demonstrated superiorization for tomography when total variation is used as the optimization criterion. In particular, we illustrated on a particular tomography problem that, in spite of its generality, superiorization produced a reconstruction that is as good as (from the points of view of constraints-compatibility and TV-minimization) what was obtained by the ASD-POCS algorithm that was specially designed for TV-minimization in tomography.

ACKNOWLEDGMENTS

The detailed and penetrating comments of three reviewers and the editors helped us to improve this paper in a significant way. The authors thank Professor Xiaochuan Pan and his co-workers from the University of Chicago for providing them with the reconstruction from their data using their implementation of their ASD-POCS algorithm. Our work is supported by the National Science Foundation Award No. DMS-1114901, the United States-Israel Binational Science Foundation (BSF) Grant No. 200912, the U.S. Department of Defense Prostate Cancer Research Program Award No. W81XWH-12-1-0122, and the U.S. Department of Army Award No. W81XWH-10-1-0170.

APPENDIX: MATHEMATICAL PROOFS

1. Conditions for strong perturbation resilience

Theorem 1. Let **P** be an algorithm for a problem structure $\langle \mathbb{T}, \mathcal{P}r \rangle$ such that, for all $T \in \mathbb{T}$, **P** is boundedly convergent for T, $\mathcal{P}r_T : \Omega \to \mathbb{R}$ is uniformly continuous, and $\mathbf{P}_T : \Delta \to \Omega$ is nonexpansive. Then **P** is strongly perturbation resilient.

Proof. We first show that there exists an $\varepsilon \in \mathbb{R}_+$ such that $O(T, \varepsilon, ((\mathbf{P}_T)^k \boldsymbol{x})_{k=0}^{\infty})$ is defined for every $\boldsymbol{x} \in \Omega$. Under the assumptions of the theorem, let $\gamma \in \mathbb{R}_+$ be such that $\mathcal{P}r_T(\boldsymbol{y}(\boldsymbol{x})) \le \gamma$, for every $\boldsymbol{x} \in \Omega$. We prove that $O(T, 2\gamma, ((\mathbf{P}_T)^k \boldsymbol{x})_{k=0}^{\infty})$ is defined for every $\boldsymbol{x} \in \Omega$ as follows. Select a particular $\boldsymbol{x} \in \Omega$. By uniform continuity of $\mathcal{P}r_T$, there exists a $\delta > 0$, such that $|\mathcal{P}r_T(\boldsymbol{z}) - \mathcal{P}r_T(\boldsymbol{y}(\boldsymbol{x}))| \le \gamma$, for any $\boldsymbol{z} \in \Omega$ for which $\|\boldsymbol{z} - \boldsymbol{y}(\boldsymbol{x})\| \le \delta$. Since **P** is convergent for T, there exists a non-negative integer K, such that $\|(\mathbf{P}_T)^K \boldsymbol{x} - \boldsymbol{y}(\boldsymbol{x})\| \le \delta$. It follows that

$$|\mathcal{P}r_T((\mathbf{P}_T)^K \boldsymbol{x})| \le |\mathcal{P}r_T((\mathbf{P}_T)^K \boldsymbol{x}) - \mathcal{P}r_T(\boldsymbol{y}(\boldsymbol{x}))| + |\mathcal{P}r_T(\boldsymbol{y}(\boldsymbol{x}))| \le 2\gamma. \tag{A1}$$

Now let $T \in \mathbb{T}$ and $\varepsilon \in \mathbb{R}_+$ be such that $O(T, \varepsilon, ((\mathbf{P}_T)^k \boldsymbol{x})_{k=0}^{\infty})$ is defined for every $\boldsymbol{x} \in \Omega$. To prove the theorem, we need to show that $O(T, \varepsilon', R)$ is defined for every $\varepsilon' > \varepsilon$ and for every sequence $R = (\boldsymbol{x}^k)_{k=0}^{\infty}$ of points in Ω for which, for all $k \ge 0$, Eq. (6) is satisfied for bounded perturbations $\beta_k \boldsymbol{v}^k$. Let ε' and R satisfy the conditions of the previous sentence.

For $k \ge 0$, we have, due to the nonexpansiveness of $\mathbf{P}_T$, that

$$\|\boldsymbol{x}^{k+1} - \mathbf{P}_T \boldsymbol{x}^k\| = \|\mathbf{P}_T(\boldsymbol{x}^k + \beta_k \boldsymbol{v}^k) - \mathbf{P}_T \boldsymbol{x}^k\| \le \|\beta_k \boldsymbol{v}^k\|. \tag{A2}$$

Denote $\|\beta_k \boldsymbol{v}^k\|$ by r_k. Clearly, $r_k \in \mathbb{R}_+$ and it follows from the definition of bounded perturbations that $\sum_{k=0}^{\infty} r_k < \infty$.

We next prove by induction that, for every pair of non-negative integers k and i,

$$\|\boldsymbol{x}^{k+i} - (\mathbf{P}_T)^i \boldsymbol{x}^k\| \le \sum_{j=k}^{k+i-1} r_j. \tag{A3}$$

Let k be an arbitrary non-negative integer. If $i = 0$, then the value is zero on both sides of the inequality and hence Eq. (A3) holds. Now assume that Eq. (A3) holds for an integer $i \geq 0$. Then, by Eq. (A2) and the nonexpansiveness of $\mathbf{P}_T$,

$$\begin{aligned}\|\boldsymbol{x}^{k+i+1} - (\mathbf{P}_T)^{i+1}\boldsymbol{x}^k\| &\leq \|\boldsymbol{x}^{k+i+1} - \mathbf{P}_T\boldsymbol{x}^{k+i}\| \\ &\quad + \|\mathbf{P}_T\boldsymbol{x}^{k+i} - (\mathbf{P}_T)^{i+1}\boldsymbol{x}^k\| \\ &\leq r_{k+i} + \|\boldsymbol{x}^{k+i} - (\mathbf{P}_T)^{i}\boldsymbol{x}^k\| \\ &\leq r_{k+i} + \sum_{j=k}^{k+i-1} r_j \\ &= \sum_{j=k}^{k+i} r_j, \end{aligned} \tag{A4}$$

which completes our inductive proof. A consequence of Eq. (A3) is that, for every pair of non-negative integers k and i,

$$\|\boldsymbol{x}^{k+i} - (\mathbf{P}_T)^{i}\boldsymbol{x}^k\| \leq \sum_{j=k}^{\infty} r_j. \tag{A5}$$

Due to the summability of the non-negative sequence $(r_k)_{k=0}^{\infty}$, the right-hand side (and hence the left-hand side) of this inequality gets arbitrarily close to zero as k increases.

Since $\mathcal{P}r_T$ is uniformly continuous, there exists a δ such that, for all $\boldsymbol{x}, \boldsymbol{y} \in \Omega$, $|\mathcal{P}r_T(\boldsymbol{x}) - \mathcal{P}r_T(\boldsymbol{y})| \leq \varepsilon' - \varepsilon$ provided that $\|\boldsymbol{x} - \boldsymbol{y}\| \leq \delta$. Select a k so that $\sum_{j=k}^{\infty} r_j \leq \delta$. By the assumption that $O(T, \varepsilon, ((\mathbf{P}_T)^k\boldsymbol{x})_{k=0}^{\infty})$ is defined for every $\boldsymbol{x} \in \Omega$, there exists a non-negative integer i for which $\mathcal{P}r((\mathbf{P}_T)^i\boldsymbol{x}^k) \leq \varepsilon$. From Eq. (A5) we have, for this k and i, that $\|\boldsymbol{x}^{k+i} - (\mathbf{P}_T)^i\boldsymbol{x}^k\| \leq \delta$ and, hence,

$$\begin{aligned}|\mathcal{P}r_T(\boldsymbol{x}^{k+i})| &\leq |\mathcal{P}r_T(\boldsymbol{x}^{k+i}) - \mathcal{P}r_T((\mathbf{P}_T)^i\boldsymbol{x}^k)| \\ &\quad + |\mathcal{P}r_T((\mathbf{P}_T)^i\boldsymbol{x}^k)| \\ &\leq (\varepsilon' - \varepsilon) + \varepsilon = \varepsilon', \end{aligned} \tag{A6}$$

proving that $O(T, \varepsilon', R)$ is defined. □

2. Nonascending vectors for convex functions

Theorem 2: Let $\phi : \mathbb{R}^J \to \mathbb{R}$ be a convex function and let $\boldsymbol{x} \in \mathbb{R}^J$. Let $\boldsymbol{g} \in \mathbb{R}^J$ satisfy the property: For $1 \leq j \leq J$, if the jth component g_j of $\boldsymbol{g}$ is not zero, then the partial derivative $\frac{\partial\phi}{\partial x_j}(\boldsymbol{x})$ of ϕ at $\boldsymbol{x}$ exists and its value is g_j. Define $\boldsymbol{d}$ to be the zero vector if $\|\boldsymbol{g}\| = 0$ and to be $-\boldsymbol{g}/\|\boldsymbol{g}\|$ otherwise. Then $\boldsymbol{d}$ is a nonascending vector for ϕ at $\boldsymbol{x}$.

Proof: The theorem is trivially true if $\|\boldsymbol{g}\| = 0$, so we assume that this is not the case. We denote by I the nonempty set of those indices j for which $g_j \neq 0$.

For $1 \leq j \leq J$, let s_j be $g_j/|g_j|$ for $j \in I$ and be 0 otherwise, and let $\boldsymbol{e}^j \in \mathbb{R}^J$ be the vector all of whose components are zero except for the jth, which is one. Then, for $1 \leq j \leq J$, there exists a $\delta_j > 0$ such that, for $0 \leq \lambda_j \leq \delta_j$,

$$\phi(\boldsymbol{x} - \lambda_j s_j \boldsymbol{e}^j) \leq \phi(\boldsymbol{x}). \tag{A7}$$

This is obvious if $s_j = 0$. Otherwise, $\frac{\partial\phi}{\partial x_j}(\boldsymbol{x})$ exists and indicates ϕ increases at $\boldsymbol{x}$ if $s_j = 1$ or that ϕ decreases at $\boldsymbol{x}$ if $s_j = -1$. The existence of the desired δ_j can be derived from the standard definition of the partial derivative as a limit.

We define $\delta > 0$ by

$$\delta = \frac{\|\boldsymbol{g}\|}{J} \min_{j \in I}\left\{\frac{\delta_j}{|g_j|}\right\}. \tag{A8}$$

Then we have that, for $0 \leq \lambda \leq \delta$,

$$\begin{aligned}\phi(\boldsymbol{x} + \lambda\boldsymbol{d}) &= \phi\left(\boldsymbol{x} - \lambda\sum_{j=1}^{J}\frac{|g_j|}{\|\boldsymbol{g}\|}s_j\boldsymbol{e}^j\right) \\ &= \phi\left(\sum_{j=1}^{J}\frac{1}{J}\left(\boldsymbol{x} - \lambda J\frac{|g_j|}{\|\boldsymbol{g}\|}s_j\boldsymbol{e}^j\right)\right) \\ &\leq \frac{1}{J}\sum_{j=1}^{J}\phi\left(\boldsymbol{x} - \lambda J\frac{|g_j|}{\|\boldsymbol{g}\|}s_j\boldsymbol{e}^j\right) \\ &\leq \frac{1}{J}\sum_{j=1}^{J}\phi(\boldsymbol{x}) \\ &= \phi(\boldsymbol{x}). \end{aligned} \tag{A9}$$

The first inequality above follows from the convexity of ϕ and the second one follows from Eq. (A7), with λ_j defined to be $\lambda J\frac{|g_j|}{\|\boldsymbol{g}\|}$, combined with Eq. (A8). Thus $\boldsymbol{d}$ is a nonascending vector for ϕ at $\boldsymbol{x}$. □

[a]Author to whom correspondence should be addressed. Electronic mail: gaborther man@yahoo.com; URL: http://www.dig.cs.gc.cuny.edu/gabor/index.html.

[1]J. O. Deasy, "Multiple local minima in radiotherapy optimization problems with dose-volume constraints," Med. Phys. **24**, 1157–1161 (1997).
[2]G. A. Ezzell, "Genetic and geometric optimization of three-dimensional radiation therapy treatment planning," Med. Phys. **23**, 293–305 (1996).
[3]A. Gustafsson, B. K. Lind, and A. Brahme, "A generalized pencil beam algorithm for optimization of radiation-therapy," Med. Phys. **21**, 343–357 (1994).
[4]A. Gustafsson, B. K. Lind, R. Svensson, and A. Brahme, "Simultaneous-optimization of dynamic multileaf collimation and scanning patterns or compensation filters using a generalized pencil beam algorithm," Med. Phys. **22**, 1141–1156 (1995).
[5]E. Lessard and J. Pouliot, "Inverse planning anatomy-based dose optimization for hdr-brachytherapy of the prostate using fast simulated annealing algorithm and dedicated objective function," Med. Phys. **28**, 773–779 (2001).
[6]R. Manzke, M. Grass, T. Nielsen, G. Shechter, and D. Hawkes, "Adaptive temporal resolution optimization in helical cardiac cone beam CT reconstruction," Med. Phys. **30**, 3072–3080 (2003).
[7]A. B. Pugachev, A. L. Boyer, and L. Xing, "Beam orientation optimization in intensity-modulated radiation treatment planning," Med. Phys. **27**, 1238–1245 (2000).
[8]D. M. Shepard, M. A. Earl, X. A. Li, S. Naqvi, and C. Yu, "Direct aperture optimization: A turnkey solution for step-and-shoot IMRT," Med. Phys. **29**, 1007–1018 (2002).
[9]C. Studholme, D. L. G. Hill, and D. J. Hawkes, "Automated three-dimensional registration of magnetic resonance and positron emission tomography brain images by multiresolution optimization of voxel similarity measures," Med. Phys. **24**, 25–35 (1997).
[10]Q. W. Wu and R. Mohan, "Algorithms and functionality of an intensity modulated radiotherapy optimization system," Med. Phys. **27**, 701–711 (2000).
[11]Y. Yu and M. C. Schell, "A genetic algorithm for the optimization of prostate implants," Med. Phys. **23**, 2085–2091 (1996).

[12]T. Z. Zhang, R. Jeraj, H. Keller, W. G. Lu, G. H. Olivera, T. R. McNutt, T. R. Mackie, and B. Paliwal, "Treatment plan optimization incorporating respiratory motion," Med. Phys. **31**, 1576–1586 (2004).
[13]M. Abdoli, M. R. Ay, A. Ahmadian, R. A. Dierckx, and H. Zaidi, "Reduction of dental filling metallic artifacts in CT-based attenuation correction of PET data using weighted virtual sinograms optimized by a genetic algorithm," Med. Phys. **37**, 6166–6177 (2010).
[14]S. Bartolac, S. Graham, J. Siewerdsen, and D. Jaffray, "Fluence field optimization for noise and dose objectives in CT," Med. Phys. **38**, S2–S17 (2011).
[15]W. Chen, D. Craft, T. M. Madden, K. Zhang, H. M. Kooy, and G. T. Herman, "A fast optimization algorithm for multicriteria intensity modulated proton therapy planning," Med. Phys. **37**, 4938–4945 (2010).
[16]J. Fiege, B. McCurdy, P. Potrebko, H. Champion, and A. Cull, "PARETO: A novel evolutionary optimization approach to multiobjective IMRT planning," Med. Phys. **38**, 5217–5229 (2011).
[17]A. Fredriksson, A. Forsgren, and B. Hardemark, "Minimax optimization for handling range and setup uncertainties in proton therapy," Med. Phys. **38**, 1672–1684 (2011).
[18]C. Holdsworth, M. Kim, J. Liao, and M. H. Phillips, "A hierarchical evolutionary algorithm for multiobjective optimization in IMRT," Med. Phys. **37**, 4986–4997 (2010).
[19]C. Holdsworth, R. D. Stewart, M. Kim, J. Liao, and M. H. Phillips, "Investigation of effective decision criteria for multiobjective optimization in IMRT," Med. Phys. **38**, 2964–2974 (2011).
[20]T. Kim, L. Zhu, T.-S. Suh, S. Geneser, B. Meng, and L. Xing, "Inverse planning for IMRT with nonuniform beam profiles using total-variation regularization (TVR)," Med. Phys. **38**, 57–66 (2011).
[21]C. Men, H. E. Romeijn, X. Jia, and S. B. Jiang, "Ultrafast treatment plan optimization for volumetric modulated arc therapy (VMAT)," Med. Phys. **37**, 5787–5791 (2010).
[22]S. N. Penfold, R. W. Schulte, Y. Censor, and A. B. Rosenfeld, "Total variation superiorization schemes in proton computed tomography image reconstruction," Med. Phys. **37**, 5887–5895 (2010).
[23]E. Y. Sidky, Y. Duchin, X. Pan, and C. Ullberg, "A constrained, total-variation minimization algorithm for low-intensity x-ray CT," Med. Phys. **38**, S117–S125 (2011).
[24]H. Stabenau, L. Rivera, E. Yorke, J. Yang, R. Lu, R. J. Radke, and A. Jackson, "Reduced order constrained optimization (ROCO): Clinical application to lung IMRT," Med. Phys. **38**, 2731–2741 (2011).
[25]Y. Yang and M. J. Rivard, "Dosimetric optimization of a conical breast brachytherapy applicator for improved skin dose sparing," Med. Phys. **37**, 5665–5671 (2010).
[26]X. Zhang, J. Wang, and L. Xing, "Metal artifact reduction in x-ray computed tomography (CT) by constrained optimization," Med. Phys. **38**, 701–711 (2011).
[27]D. Butnariu, R. Davidi, G. T. Herman, and I. G. Kazantsev, "Stable convergence behavior under summable perturbations of a class of projection methods for convex feasibility and optimization problems," IEEE J. Sel. Top. Signal Process. **1**, 540–547 (2007).
[28]R. Davidi, G. T. Herman, and Y. Censor, "Perturbation-resilient block-iterative projection methods with application to image reconstruction from projections," Int. Trans. Oper. Res. **16**, 505–524 (2009).
[29]Y. Censor, R. Davidi, and G. T. Herman, "Perturbation resilience and superiorization of iterative algorithms," Inverse Probl. **26**, 065008 (2010).
[30]T. Nikazad, R. Davidi, and G. T. Herman, "Accelerated perturbation-resilient block-iterative projection methods with application to image reconstruction," Inverse Probl. **28**, 035005 (2012).
[31]G. T. Herman and R. Davidi, "Image reconstruction from a small number of projections," Inverse Probl. **24**, 045011 (2008).
[32]E. Garduño, R. Davidi, and G. T. Herman, "Reconstruction from a few projections by ℓ_1-minimization of the Haar transform," Inverse Probl. **27**, 055006 (2011).
[33]R. L. Rardin and R. Uzsoy, "Experimental evaluation of heuristic optimization algorithms: A tutorial," J. Heuristics **7**, 261–304 (2001).
[34]L. Wernisch, S. Hery, and S. J. Wodak, "Automatic protein design with all atom force-fields by exact and heuristic optimization," J. Mol. Biol. **301**, 713–736 (2000).
[35]S. H. Zanakis and J. R. Evans, "Heuristic optimization: Why, when, and how to use it," Interfaces **11**, 84–91 (1981).
[36]G. T. Herman and W. Chen, "A fast algorithm for solving a linear feasibility problem with application to intensity-modulated radiation therapy," Linear Algebra Appl. **428**, 1207–1217 (2008).
[37]E. S. Helou Neto and Á. R. De Pierro, "Incremental subgradients for constrained convex optimization: A unified framework and new methods," SIAM J. Optim. **20**, 1547–1572 (2009).
[38]E. S. Helou Neto and Á. R. De Pierro, "On perturbed steepest descent methods with inexact line search for bilevel convex optimization," Optim. **60**, 991–1008 (2011).
[39]E. A. Nurminski, "Envelope stepsize control for iterative algorithms based on Fejer processes with attractants," Optim. Methods Software **25**, 97–108 (2010).
[40]P. L. Combettes and J. Luo, "An adaptive level set method for nondifferentiable constrained image recovery," IEEE Trans. Image Process. **11**, 1295–1304 (2002).
[41]P. L. Combettes and J.-C. Pesquet, "Image restoration subject to a total variation constraint," IEEE Trans. Image Process. **13**, 1213–1222 (2004).
[42]E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," Phys. Med. Biol. **53**, 4777–4807 (2008).
[43]J. Bian, J. H. Siewerdsen, X. Han, E. Y. Sidky, J. L. Prince, C. A. Pelizzari, and X. Pan, "Evaluation of sparse-view reconstruction from flat-panel-detector cone-beam CT," Phys. Med. Biol. **55**, 6575–6599 (2010).
[44]M. Defrise, C. Vanhove, and X. Liu, "An algorithm for total variation regularization in high-dimensional linear problems," Inverse Probl. **27**, 065002 (2011).
[45]Y. Censor, W. Chen, P. L. Combettes, R. Davidi, and G. T. Herman, "On the effectiveness of projection methods for convex feasibility problems with linear inequality constraints," Comput. Optim. Appl. **51**, 1065–1088 (2012).
[46]J. Bioucas-Dias and M. Figueiredo, "A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration," IEEE Trans. Image Process. **16**, 2992–3004 (2007).
[47]T. Goldstein and S. Osher, "The split Bregman method for L1 regularized problems," SIAM J. Imaging Sci. **2**, 323–343 (2009).
[48]L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," IEEE Trans. Med. Imaging **1**, 113–122 (1982).
[49]E. Levitan and G. T. Herman, "A maximum *a posteriori* probability expectation maximization algorithm for image reconstruction in emission tomography," IEEE Trans. Med. Imaging **6**, 185–192 (1987).
[50]W. Jin, Y. Censor, and M. Jiang, "A heuristic superiorization-like approach to bioluminescence tomography," in *Proceedings of the International Federation for Medical and Biological Engineering (IFMBE)* (Springer-Verlag, Berlin, 2012), Vol. 39, pp. 1026–1029.
[51]H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," IEEE Trans. Med. Imaging **13**, 601–609 (1994).
[52]T. Elfving, "Block-iterative methods for consistent and inconsistent linear equations," Numer. Math. **35**, 1–12 (1980).
[53]P. P. B. Eggermont, G. T. Herman, and A. Lent, "Iterative algorithms for large partitioned linear systems, with applications to image reconstruction," Linear Algebra Appl. **40**, 37–67 (1981).
[54]R. Aharoni and Y. Censor, "Block-iterative projection methods for parallel computation of solutions to convex feasibility problems," Linear Algebra Appl. **120**, 165–175 (1989).
[55]G. T. Herman, *Fundamentals of Computerized Tomography: Image Reconstruction from Projections*, 2nd ed. (Springer, New York, 2009).
[56]J. F. P. J. Abascal, J. Chamorro-Servent, J. Aguirre, S. Arridge, T. Correia, J. Ripoli, J. J. Vaquero, and M. Desco, "Fluorescence diffuse optical tomography using the split Bregman method," Med. Phys. **38**, 6275–6284 (2011).
[57]R. Davidi, G. T. Herman, and J. Klukowska, SNARK09: A programming system for the reconstruction of 2D images from 1D projections, 2009 (available URL: http://www.snark09.com).
[58]G. T. Herman and L. B. Meyer, "Algebraic reconstruction techniques can be made computationally efficient," IEEE Trans. Med. Imaging **12**, 600–609 (1993).
[59]W. Zbijewski and F. J. Beekman, "Characterization and suppression of edge and aliasing artefacts in iterative x-ray CT reconstruction," Phys. Med. Biol. **49**, 145–157 (2004).

Noname manuscript No.
(will be inserted by the editor)

Projected Subgradient Minimization Versus Superiorization

Yair Censor[1], Ran Davidi[2], Gabor T. Herman[3], Reinhard W. Schulte[4] and Luba Tetruashvili[1]

February 4, 2013

Abstract The projected subgradient method for constrained minimization repeatedly interlaces subgradient steps for objective function descent with projections onto the feasible region, which is the intersection of closed convex constraints sets, to regain feasibility. The latter pose a computational difficulty and therefore the projected subgradient method is applicable only when the feasible region is "simple to project onto". In contrast with this, in the superiorization methodology a feasibility-seeking algorithm leads the overall process and objective function descent steps are interlaced into it. This makes a difference because the feasibility-seeking algorithm employs projections onto the individual constraints sets and not onto the whole feasible region.

We present the two approaches side-by-side and demonstrate their performance on a problem of computerized tomography image reconstruction posed as a constrained minimization problem aiming at finding a constraint-compatible solution that has a reduced value of the total variation of the reconstructed image.

Keywords constrained minimization, feasibility-seeking, bounded convergence, superiorization, projected subgradient method, proximity function, strong perturbation resilience, image reconstruction, computerized tomography

PACS 65K05, 90C59, 65B99, 49M30, 90C90, 90C30

1 Introduction

Our aim in this paper is to expose the recently-developed superiorization methodology (SM) and its ideas to the optimization community by "confronting" it with the projected subgradient method (PSM). The reasons for this choice are explained below. Throughout this paper we assume that Ω is a nonempty subset of the J-dimensional

[1]Department of Mathematics, University of Haifa, Mt. Carmel, 31905 Haifa, Israel
[2]Department of Radiation Oncology, Stanford University, Stanford, CA 94305, USA
[3]Department of Computer Science, The Graduate Center , City University of New York, New York, NY 10016, USA
[4]Department of Radiation Medicine, Loma Linda University Medical Center, Loma Linda, CA 92354 USA

Euclidean space R^J. We consider constrained minimization problems of the form

$$\text{minimize}\,\{\phi(x) \mid x \in C\}\,, \tag{1}$$

where $\phi : R^J \to R$ is an objective function and $C \subseteq \Omega$ is a given feasible set. Such nonlinear constrained optimization problems lie at the heart of optimization theory and practice and constitute mathematical models for many scientific and real-world applications. Various theories and different methods abound for fulfilling the task of finding a point that is both feasible for the constraints and renders a minimal value to the objective function. Some approaches, known as regularization methods, fold the feasibility-seeking part of the overall task into the minimization via penalty functions. Others assume a feasible point as a starting point and conduct the search for minimality while preventing the iterates of an algorithm from leaving the constraints by using barrier functions. Still others deal with feasibility-seeking and the search for minimality as two separate tasks, see, e.g., [19].

In this paper we juxtapose the *projected subgradient method* (PSM) [sometimes referred to also as the *(sub)gradient projection method*] with the recently-developed *superiorization methodology* (SM) and demonstrate their performance on a large-size real-world application that is modeled, and needs to be solved, as a constrained minimization problem. It is not claimed that PSM is the best optimization method for solving (1) and there are many different alternative methods to which SM could be compared. So why did we chose to confront PSM with our SM? In a nutshell, our answer is that both methods interlace objective-function-reduction steps with steps oriented toward feasibility, but they differ in how they restore or preserve feasibility. We now outline these two methods and explain our choice in more detail.

The PSM for constrained minimization has been extensively investigated, see, e.g., [41, Subsection 7.1.2], [27, Subsection 3.2.3]. Its roots are in the work of Shor [42] for the unconstrained case and in the work of Polyak [38,39] for the constrained case. More recent work can be found in, e.g., [5]. In order to apply the PSM to solving (1) we need to assume that C is a nonempty closed convex set and that ϕ is a convex function. PSM generates a sequence of iterates $\left\{x^k\right\}_{k=0}^{\infty}$ according to the recursion formula

$$x^{k+1} = P_C\left(x^k - t_k\phi'\left(x^k\right)\right), \tag{2}$$

where $t_k > 0$ is a step-size, $\phi'\left(x^k\right) \in \partial\phi\left(x^k\right)$ is a subgradient of ϕ at x^k, and P_C stands for the orthogonal (least Euclidean norm) projection onto the set C. The underlying philosophy is to perform unconstrained objective function descent steps via $q^k := x^k - t_k\phi'\left(x^k\right)$ and regain feasibility with respect to C after each such step by projecting q^k onto C. Many published studies give various sets of conditions under which sequences generated by an algorithm that includes an iterative step like (2) converge or have certain other desirable properties and a great deal of experimental work has been done and published about applying such methods computationally.

A major difficulty with (2) is the need to perform, within each iterative step, the orthogonal projection. If the feasible set C is not "simple to project onto" then the projection requires an independent inner-loop calculation to minimize the distance from the point q^k to the set C, which can be costly and hamper the overall effectiveness of an algorithm that uses (2). Also, if the inner loop converges to the projection onto C only in the limit, then in practical implementations it will have to be stopped after

a finite number of steps and so x^{k+1} will be only an approximation to the projection onto C and it could even happen that it is not in C.

Even if we set aside our worries about projecting onto C in (2), there are still two concerns when applying PSM to real-world problems. One is that the iterative process usually converges to the desired solution only in the limit. In practice, some stopping rule is applied to terminate the process and the output at that time may not even be in C and, even if it is in C, it is most unlikely to be the minimizer of ϕ over C. The second problem in real-world applications comes from the fact that the constraints, derived from the real-world problem, may not be consistent (e.g., because they come from noisy measurements) and so C is empty.

Both of these objections can be handled by replacing the notion of a fixed feasible set C by that of a nonnegative real-valued proximity function $Prox_C : \Omega \to R_+$. This function serves as an indicator of how incompatible a vector x is with the constraints. In such a formulation the merit of the actual output x of any algorithm is indicated by the smallness of the two numbers $Prox_C(x)$ and $\phi(x)$. For the formulation of (1), we would define $Prox_C$ so that its range is the ray of nonnegative real numbers with $Prox_C(x) = 0$ if, and only if, $x \in C$ and then the constrained minimization problem (1) is precisely that of finding an x that is a minimizer of $\phi(x)$ over $\{x \mid Prox_C(x) = 0\}$. The above discussion allows us to do away with the nonemptiness assumption and also to compare the merits of actual outputs of algorithms that only approximate the aim of the constrained minimization problem.

The recently invented SM incorporates the ideas of the previous paragraph in its very foundation and formulates the problem with the function $Prox_C$ instead of the set C. The underlying idea of SM is that many iterative algorithms that produce outputs x for which $Prox_C(x)$ is small are *strongly perturbation resilient* in the sense that, even if certain kinds of changes are made at the end of each iterative step, the algorithm still produces an output x' for which $Prox_C(x')$ is not larger. This property is exploited by using permitted changes to steer the algorithm to an output that has not only a small $Prox_C$ value, but has also a small ϕ value. The algorithm that incorporates such a steering process is referred to as the *superiorized version* of the original iterative algorithm. The main practical contribution of SM is the automatic creation of the superiorized version, according to a given objective function ϕ, of just about any iterative algorithm that aims at producing an x for which $Prox_C(x)$ is small.

Nevertheless, in order to carry out our comparative study, we restrict our attention here to a subset of all possible problems to which not only SM but also PSM is applicable. We assume that we are given a family of constraints $\{C_\ell\}_{\ell=1}^{L}$, where each set C_ℓ is a nonempty closed convex subset of R^J such that

$$C = \bigcap_{\ell=1}^{L} C_\ell \tag{3}$$

is a nonempty subset of Ω and that it is the feasible set C of (1). Under these assumptions, we illustrate the SM by the superiorization of feasibility-seeking *projection method*s, see, e.g., [1, 2, 3, 9, 15] and the recent monograph [8]. Such methods use projections onto the individual sets C_ℓ in order to generate a sequence $\left\{x^k\right\}_{k=0}^{\infty}$ that converges to a point $x^* \in C$. Therefore, contrary to the PSM method, one does not need to assume that C is a "simple to project onto" set, but rather that the individual sets C_ℓ have this property. The latter is indeed often the case, such as, for example,

when the sets C_ℓ are hyperplanes or half-spaces onto which we can project easily, but their intersection is not "simple to project onto".

The SM is accurately presented below in Section 4. But the discussion above is sufficient to explain why we chose the PSM method and the SM for our comparative study. Namely, both methods interlace objective-function-reduction steps with steps oriented toward feasibility. But exactly here lies a big difference between the two approaches. The PSM method requires that feasibility is regained after subgradient nonascent steps by performing a projection onto C, whereas in the SM the feasibility-seeking projection method proceeds by projecting (in a well-defined algorithmically-structured regime dictated by the specific projection method) onto the individual sets C_i and not onto the whole feasible set C. This has a potentially great computational advantage.

In this paper we demonstrate the approaches of SM and PSM on a realistically-large-size problem with data that arise from the significant problem of x-ray computed tomography (CT) with total variation (TV) minimization. In Section 2 we discuss some related work, in Section 3 we describe the PSM method that we use, in Section 4 the SM is presented, and in Section 5 the computational demonstration is reported, followed by some conclusions in Section 6.

2 Related previous work

In this section we briefly describe the relationships of superiorization with previously published work on related methods. The points made in the second half of Subsection 2.2 are of particular importance.

2.1 Previous work on superiorization

The superiorization methodology was first proposed (although without using the term superiorization) in [7]. In that work perturbation resilience (without using this term) was proved for the general class of *string-averaging projection* (SAP) methods, see [11, 12, 13, 14, 36], that use orthogonal projections and relate to consistent constraints. Subsequent investigations and developments were done in [10, 17, 21, 31, 37]. In [10] the methodology was formulated over general *problem structures* which enabled rigorous analysis and revealed that the approach is not limited to feasibility and optimization. In [17] perturbation resilience was analyzed for the class of *block-iterative projection* (BIP) methods, see [1, 2, 3, 9, 15], and applied in this manner. In [21] the advantages of superiorization for image reconstruction from a small number of projections was studied, and in [31] two acceleration schemes based on (symmetric and nonsymmetric) BIP methods were proposed and experimented with. In [37] total variation superiorization schemes in proton computed tomography (pCT) image reconstruction were investigated.

In [22] we introduced the notion of ε-compatibility into the superiorization approach in order to handle inconsistent constraints. This enabled us to close the logical discrepancy between the assumption of consistency of constraints and the actual experimental work done previously. We also introduced there the new notion of strong perturbation resilience which generalizes the previously used notion of perturbation resilience. Algorithmically, the new superiorized algorithm introduced there (and used

here) is different from all previous ones in that it uses the notion of nonascending direction and in that it allows several perturbation steps for each feasibility-seeking step, an aspect that has practical advantages.

In [23] superiorization was applied to the *expectation maximization* (EM) algorithm instead of the feasibility-seeking projection methods that were used in superiorization previously. The approach was implemented there to solve an inverse problem of *bioluminescence tomography* (BLT) image reconstruction. Such EM superiorization was investigated further and applied to a problem of *Single Photon Emission Computed Tomography* (SPECT) in [25]. Most recently, the SM was further investigated numerically, along with many projection methods for the feasibility problem and for the best approximation problem, in [4].

2.2 The works of of Nurminski, of Helou Neto and De Pierro, and of Nedić

Our superiorization methodology should be distinguished from the works of Helou Neto and De Pierro [30,29], of Nedić [40,26], and of Nurminski [33,32,34,35]. The lack of cross-referencing between some of these papers shows that, in spite of the similarities between their approaches, their results were apparently reached independently.

Nurminski: The algorithms of Nurminski use Fejér operators, that can be used in feasibility-seeking, and introduces into them disturbances with diminishing step-sizes $\lambda_k \to 0$ as $k \to \infty$, where the rate of this tendency is such that $\sum_{k=0}^{\infty} \lambda_k = +\infty$. Under these conditions and a variety of additional assumptions, Nurminski showed asymptotic convergence of the iterates generated by his algorithms to a minimum point of the constrained minimization problem.

Helou Neto and De Pierro: The framework proposed by Helou Neto and De Pierro uses interlacing of "feasibility operators" with "optimality operators" with the aim of creating exact constrained minimization algorithms. Similarly to Nurminski, they employ diminishing step-sizes $\lambda_k \to 0$ as $k \to \infty$ such that $\sum_{k=0}^{\infty} \lambda_k = +\infty$. Under these conditions and a variety of additional assumptions, different than those of Nurminski, they show asymptotic convergence of the iterates generated by their algorithmic framework to a minimum point of the constrained minimization problem.

However, when it comes to derivation of specific algorithms from the general framework of [29, Equation (3)], their *feasibility operator* $\mathcal{F}$ invariably takes the form

$$\mathcal{F}_F(x) = x - \mu(x)\nabla F(x), \tag{4}$$

where the function $F(x)$, whose gradient is calculated, is "a convex function such that the set of minima of this function coincides with the set $\mathcal{X}$ [in [29, Equation (3)] $\mathcal{X}$ is the feasible set of the minimization problem and should be identified with $C = \bigcap_{\ell=1}^{L} C_\ell$ in our notation] when it is not empty and defines a solution in an appropriate way (least squares for example) otherwise" and $\mu(x)$ are some parameters that are restricted in a particular manner as in [29, Lemma 7 or Corollary 8]. This function $F(x)$ may be identified with what we call the proximity function $Prox_C(x)$ in the superiorization methodology, but our feasibility-seeking algorithms, and hence their versions produced by our superiorization methodology, are not limited to the form of (4).

Nedić: The overall approach of Nedić is to apply gradient and subgradient iterative methods for the objective function minimization and interlace into them random feasibility updates. Her resulting "random projection method" [26, Equation (4)] bears structural similarity to our SM but again the diminishing step-sizes $\alpha_k \to 0$ as $k \to \infty$

in [26, Equation (4)] must be such that $\sum_{k=0}^{\infty} \alpha_k = +\infty$, see [26, Proposirion 1 and Proposition 2]. The randomness refers to the way the constraints are picked up for the feasibility updates.

To summarize, there are various differences among all the above works and between them and our work, differences in overall setup of the problems, differences in the assumptions used for the various convergence results, etc. This is not the place for a full review of all these differences. But we wish to clarify the fundamental difference between them and the SM. The point is that when two activities are interlaced (here, feasibility steps and objective function reduction steps) then once the process is running all such methods look similar. From looking at the iterative formulas one cannot tell if (a) "feasibility steps are interlaced into an iterative gradient scheme for objective function minimization" or if (b) "objective function reduction steps are interlaced into an iterative projections scheme for feasibility-seeking".

The common thread of all above mentioned works is that they fall into the category (a), while the SM is of the kind (b). In all methods of category (a) the condition that is needed to guarantee convergence to a constrained minimum point is that the diminishing step-sizes $\alpha_k \to 0$ as $k \to \infty$ must be such that $\sum_{k=0}^{\infty} \alpha_k = +\infty$. In contrast, since the feasibility-seeking projection method is the "leader" of the overall process in the SM, we must have that the perturbations (that do the objective function reduction) will use diminishing step-sizes $\beta_k \to 0$ as $k \to \infty$ but such that $\sum_{k=0}^{\infty} \beta_k < \infty$. The latter condition guarantees the perturbation resilience of the original feasibility-seeking projection method so that, regardless of the interlaced objective function reduction steps, the overall process converges to a feasible, or ε-compatible, point of the constraints.

Yet another fundamental difference between the superiorization methodology and the algorithms of category (a) mentioned above is that those algorithms perform the interlaced objective function descent and feasibility steps alternatingly according to a rigid predetermined scheme, whereas in the superiorization methodology the activation of these steps and the decisions whether to keep an iterate or discard it are done inside the superiorized algorithm in a controlled and automatically-supervised manner. Thus, the superiorization methodology has the following features not present in the algorithms of category (a) mentioned above: (i) it conducts iterations of a feasibility-seeking projection method which is strongly perturbation resilient (as defined below), (ii) it interlaces objective function nonascent steps into the process in a controlled and automatically-supervised manner, (iii) it is not known to guarantee convergence to a solution of the constrained minimization problem and it might (we do not know if this is so or not) instead, only be shown to lead to a feasible point whose objective function value is less than that of a feasible point that would have been reached by the same feasibility-seeking projection method without the perturbations exercised by the superiorized algorithm.

2.3 Adaptive steepest descent projections onto convex sets

The *adaptive steepest descent projections onto convex sets* (ASD-POCS) algorithm is described in detail in [44] and its use has since been reported in a number of subsequent publications (for example, in [43]) in the CT literature. It interlaces function descent steps with projections onto convex sets. However, it is not as general as the SM; see [22] for a comparison.

3 The projected subgradient method

In this section we describe the projected subgradient method (PSM) and how we apply it. We follow the presentation given in [27, Subsection 3.2.3], using, in particular, Theorem 3.2.2 therein. This convergence theorem requires convexity and local Lipschitz continuity of the objective function ϕ and that the feasible set C in (1) is closed and convex. The latter is indeed the case here since we assume that the feasible set C is given by (3), with all the conditions stated there. The PSM generates a sequence of iterates $\left\{x^k\right\}_{k=0}^{\infty}$ according to the recursion formula

$$x^{k+1} = P_C\left(x^k - t_k \phi'\left(x^k\right)\right), \tag{5}$$

where $t_k > 0$ is a step-size, $\phi'\left(x^k\right) \in \partial\phi\left(x^k\right)$ is a subgradient of ϕ at x^k, and P_C stands for the orthogonal projection onto the set C. The PSM has been investigated under several types of step-size rules. Some commonly used such rules can be found, e.g., in [6]. We adopt a nonsummable diminishing step-length rule of the form $t_k = \gamma_k / \left\|\phi'\left(x^k\right)\right\|$, where $\gamma_k \geq 0$, $\lim_{k\to\infty}\gamma_k = 0$, $\sum_{k=0}^{\infty}\gamma_k = \infty$. All the above is precisely identical with the method in [27, Equation (3.2.8)] and under these conditions [27, Theorem 3.2.2] guarantees convergence of $\left\{\phi\left(x^k\right)\right\}_{k=0}^{\infty}$ to a minimum value of ϕ over C.

The next issue that we discuss is how to compute the projection P_C which is required in the PSM (5). To do this, one has to solve the optimization problem of finding

$$\arg\min_x \left\{\frac{1}{2}\|x-q\|^2 \mid x \in C = \bigcap_{\ell=1}^{L} C_\ell\right\}, \tag{6}$$

for $q = x^k - t_k\phi'\left(x^k\right)$ that has been calculated. We adopt the well-trodden path of solving (6) via duality theory. We introduce new vector variables $y^\ell \in R^J$, for all $\ell = 1, 2, \ldots, L$, and rewrite the problem (6) as the equivalent one of finding

$$\begin{cases} \arg\min\limits_{x,y^1,y^2,\ldots,y^L} \sum_{\ell=1}^{L} \frac{1}{2}\left\|y^\ell - q\right\|^2 + \frac{1}{2}\|x-q\|^2, \\ \text{such that} \quad y^\ell = x, \quad \ell = 1, 2, \ldots, L \\ \text{and} \quad\quad\; y^\ell \in C_\ell, \; i = 1, 2, \ldots, L. \end{cases} \tag{7}$$

By assigning dual vectors $\lambda^\ell \in R^J$, $\ell = 1, 2, \ldots, L$, to every equality constraint in (7) we obtain the dual function

$$\begin{aligned} D\left(\lambda^1, \lambda^2, \ldots, \lambda^L\right) := & \sum_{\ell=1}^{L} \min\left\{\frac{1}{2}\left\|y^\ell - q\right\|^2 + \left\langle\lambda^\ell, y^\ell\right\rangle \mid y^\ell \in C_\ell\right\} \\ & + \min\left\{\frac{1}{2}\|x-q\|^2 - \left\langle\sum_{\ell=1}^{L}\lambda^\ell, x\right\rangle \mid x \in R^J\right\}. \end{aligned} \tag{8}$$

Some rearrangements of terms then lead to

$$\begin{aligned} D\left(\lambda^1, \lambda^2, \ldots, \lambda^L\right) = & \sum_{\ell=1}^{L} \min\left\{\frac{1}{2}\left\|y^\ell - \left(q - \lambda^\ell\right)\right\|^2 - \frac{1}{2}\left\|q - \lambda^\ell\right\|^2 \mid y^\ell \in C_\ell\right\} \\ & + \min\left\{\frac{1}{2}\|x-q\|^2 - \left\langle\sum_{\ell=1}^{L}\lambda^\ell, x\right\rangle \mid x \in R^J\right\}. \end{aligned} \tag{9}$$

By solving the minimization problems on the right-hand side of the above equation we obtain

$$\begin{aligned} D\left(\lambda^1, \lambda^2, \ldots, \lambda^L\right) = \sum_{\ell=1}^{L} \left(\frac{1}{2} \left\| \left(q - \lambda^\ell\right) - P_{C_\ell}\left(q - \lambda^\ell\right) \right\|^2 - \frac{1}{2} \left\| q - \lambda^\ell \right\|^2 \right) \\ - \frac{1}{2} \left\| \sum_{\ell=1}^{L} \lambda^\ell \right\|^2 - \left\langle \sum_{\ell=1}^{L} \lambda^\ell, q \right\rangle. \end{aligned} \tag{10}$$

It can be shown that this function is concave and has a Lipschitz continuous gradient. When the projections onto each C_ℓ are available, then the values and gradients of this dual function D are also available. Thus, the dual problem is the convex unconstrained optimization problem of finding

$$\arg \max_{\lambda^1, \lambda^2, \ldots, \lambda^L} \left\{ D\left(\lambda^1, \lambda^2, \ldots, \lambda^L\right) \mid \lambda^\ell \in R^J, \quad \ell = 1, 2, \ldots, L \right\}, \tag{11}$$

where the objective function $D\left(\lambda^1, \lambda^2, \ldots, \lambda^L\right)$ is concave and smooth, and its values and gradients are available at any point $\left(\lambda^1, \lambda^2, \ldots, \lambda^L\right)$. The optimal point x^* of the primal problem (7) is then given by

$$x^* = q + \sum_{\ell=1}^{L} \lambda^{*\ell}, \tag{12}$$

where $\left(\lambda^{*1}, \lambda^{*2}, \ldots, \lambda^{*L}\right)$ is an optimal point of the dual problem (11).

A possible drawback of this approach via duality theory is that by introducing new variables in (7) we increase the number of variables in the dual problem. It turns out that in some special cases of the problem (6) we can efficiently solve the corresponding dual problem without enlarging its size. One of these special cases occurs when in problem (6) $L = I + 1$, the sets $C_i := \left\{ x \in R^J \mid \left\langle a^i, x \right\rangle = b_i \right\}$, for $i = 1, 2, \ldots, I$, are hyperplanes where the vector a^i is, for $i = 1, 2, \ldots, I$, the ith row of the matrix A, $b = (b_1, b_2, \ldots, b_I)$, and $C_{I+1} := \left\{ x \in R^J \mid 0 \le x \le 1 \right\}$ is an additional convex set of box constraints; that is, the problem (6) takes the form of finding

$$\arg \min_x \left\{ \frac{1}{2} \|x - q\|^2 \mid Ax = b \text{ and } 0 \le x \le 1 \right\}. \tag{13}$$

In this case, we use a dual vector $\lambda \in R^I$ and the dual problem becomes that of finding

$$\arg \max_\lambda \left\{ \begin{array}{l} \frac{1}{2} \left\| q - A^T\lambda - P_{C_{I+1}}\left(q - A^T\lambda\right) \right\|^2 \\ -\frac{1}{2} \left\| q - A^T\lambda \right\|^2 - \langle \lambda, b \rangle + \frac{1}{2} \|q\|^2 \mid \lambda \in R^I \end{array} \right\}, \tag{14}$$

where $P_{C_{I+1}}$ denotes the orthogonal projection onto the set C_{I+1}. The optimal point x^* of the problem (13) can be computed by

$$x^* = P_{C_{I+1}}\left(q - A^T\lambda^*\right), \tag{15}$$

where the vector λ^* is the optimal solution of the dual problem (14). It can be shown that the objective function $D(\lambda)$ of this dual problem (14) is convex with Lipschitz continuous gradients given by

$$\nabla D(\lambda) = A P_{C_{I+1}}\left(q - A^T\lambda\right) - b. \tag{16}$$

The problem (14) can be solved by the optimal method of Nesterov [28]. We give its description for the problem of unconstrained minimization of a function $f(x)$, assuming that $f(x)$ is convex and continuously differentiable with Lipschitz continuous gradients; i.e., that there exists a constant $L > 0$ such that, for all x, $y \in R^J$,

$$\|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\| . \tag{17}$$

Nesterov's Algorithm
(N1) **Initialization:** Select a point $y^0 \in R^J$ and put $k = 0$, $\beta_0 = 1$, $x^{-1} = y^0$, $\alpha_{-1} = \left\|y^0 - z\right\| / \left\|\nabla f\left(y^0\right) - \nabla f\left(z\right)\right\|$, where z is any point in R^J such that $z \neq y^0$ and $\nabla f\left(z\right) \neq \nabla f\left(y^0\right)$.
(N2) **Iterative Step:** Given x^{k-1}, y^k, α_{k-1} and β_k
(N2.1) Calculate the smallest index $s \geq 0$ for which the following inequality holds

$$f\left(y^k\right) - f\left(y^k - 2^{-s}\alpha_{k-1}\nabla f\left(y^k\right)\right) \geq 2^{-s-1}\alpha_{k-1}\left\|\nabla f\left(y^k\right)\right\|^2 . \tag{18}$$

(N2.2) Calculate the next iterate by

$$\alpha_k = 2^{-s}\alpha_{k-1} \text{ and } x^k = y^k - \alpha_k \nabla f\left(y^k\right), \tag{19}$$

and update

$$\beta_{k+1} = \left(1 + \frac{1}{2}\sqrt{4\beta_k^2 + 1}\right), \tag{20}$$

and

$$y^{k+1} = x^k + \frac{\beta_k - 1}{\beta_{k+1}}\left(x^k - x^{k-1}\right). \tag{21}$$

When a stopping rule applies, then the point x^k is the output of the method.

4 The superiorization methodology

In this section we present a restricted version of the SM of [22] adapted to our problem (1). As discussed in Section 1, we associate with the feasible set C in (1) a proximity function $Prox_C : \Omega \to R_+$ that is an indicator of how incompatible an $x \in \Omega$ is with the constraints. For any given $\varepsilon > 0$, a point $x \in \Omega$ for which $Prox_C(x) \leq \varepsilon$ is called an *ε-compatible solution* for C. We further assume that we have, for the C in (1), a feasibility-seeking *algorithmic operator* $\boldsymbol{A}_C : R^J \to \Omega$, with which we define the following basic algorithm.

The Basic Algorithm
(B1) **Initialization**: Choose an arbitrary $x^0 \in \Omega$,
(B2) **Iterative Step**: Given the current iterate x^k, calculate the next iterate x^{k+1} by

$$x^{k+1} = \boldsymbol{A}_C\left(x^k\right). \tag{22}$$

The following definition helps to evaluate the output of the Basic Algorithm upon termination by a stopping rule.

Definition 1 The ε-output of a sequence

Given $C \subseteq R^J$, a proximity function $Prox_C : \Omega \to R_+$, a sequence $\left\{x^k\right\}_{k=0}^{\infty} \subset \Omega$ and an $\varepsilon > 0$, then an element x^K of the sequence which has the properties: (i) $Prox_C\left(x^K\right) \leq \varepsilon$, and (ii) $Prox_C\left(x^k\right) > \varepsilon$ for all $0 \leq k < K$, is called an `$\varepsilon$-output of the sequence` $\left\{x^k\right\}_{k=0}^{\infty}$ `with respect to the pair` $(C,\ Prox_C)$. We denote it by $O\left(C, \varepsilon, \left\{x^k\right\}_{k=0}^{\infty}\right) = x^K$.

Clearly, an ε-output $O\left(C, \varepsilon, \left\{x^k\right\}_{k=0}^{\infty}\right)$ of a sequence $\left\{x^k\right\}_{k=0}^{\infty}$ might or might not exist, but if it does, then it is unique. If $\left\{x^k\right\}_{k=0}^{\infty}$ is produced by an algorithm intended for the feasible set C, such as the Basic Algorithm, without a termination criterion, then $O\left(C, \varepsilon, \left\{x^k\right\}_{k=0}^{\infty}\right)$ is the *output* produced by that algorithm when it includes the termination rule to stop when an ε-compatible solution for C is reached.

Definition 2 Strong perturbation resilience

Assume that we are given a $C \subseteq \Omega$, a proximity function $Prox_C$, an algorithmic operator $\boldsymbol{A}_C$ and an $x^0 \in \Omega$. We use $\left\{x^k\right\}_{k=0}^{\infty}$ to denote the sequence generated by the Basic Algorithm when it is initialized by x^0. The Basic Algorithm is said to be `strongly perturbation resilient` if the following hold:

(i) there exist an $\varepsilon > 0$ such that the ε-output $O\left(C, \varepsilon, \left\{x^k\right\}_{k=0}^{\infty}\right)$ is defined for every $x^0 \in \Omega$;

(ii) for every $\varepsilon > 0$, for which the ε-output $O\left(C, \varepsilon, \left\{x^k\right\}_{k=0}^{\infty}\right)$ is defined for every $x^0 \in \Omega$, we have also that the ε'-output $O\left(C, \varepsilon', \left\{y^k\right\}_{k=0}^{\infty}\right)$ is defined for every $\varepsilon' > \varepsilon$ and for every sequence $\left\{y^k\right\}_{k=0}^{\infty}$ generated by

$$y^{k+1} = \boldsymbol{A}_C\left(y^k + \beta_k v^k\right), \text{ for all } k \geq 0, \tag{23}$$

where the vector sequence $\left\{v^k\right\}_{k=0}^{\infty}$ is bounded and the scalars $\{\beta_k\}_{k=0}^{\infty}$ are such that $\beta_k \geq 0$, for all $k \geq 0$, and $\sum_{k=0}^{\infty} \beta_k < \infty$.

Definition 3 Bounded convergence

Assume that we are given a $C \subseteq R^J$, a proximity function $Prox_C$ and an algorithmic operator $\boldsymbol{A}_C : R^J \to \Omega$. Then the Basic Algorithm is said to be `convergent over` Ω if for every $x^0 \in \Omega$ there exist the limit $\lim_{k\to\infty} x^k = y\left(x^0\right)$ and $y\left(x^0\right) \in \Omega$. It is said to be `boundedly convergent over` Ω if, in addition, there exists a $\gamma \geq 0$ such that $Prox_C\left(y\left(x^0\right)\right) \leq \gamma$ for every $x^0 \in \Omega$.

The next theorem, which gives sufficient conditions for strong perturbation resilience of the Basic Algorithm, has been proved in [22, Theorem 1] (in different wording).

Theorem 1 *Assume that we are given a $C \subseteq R^J$, a proximity function $Prox_C$ and an algorithmic operator $\boldsymbol{A}_C : R^J \to \Omega$. If $\boldsymbol{A}_C$ is nonexpansive and is such that it defines a boundedly convergent Basic Algorithm and if the proximity function $Prox_C$ is uniformly continuous, then the Basic Algorithm defined by $\boldsymbol{A}_C$ is strongly perturbation resilient.*

Along with the $C \subseteq R^J$, we look at the objective function $\phi : R^J \to R$, with the convention that a point in R^J for which the value of ϕ is smaller is considered *superior* to a point in R^J for which the value of ϕ is larger. The essential idea of the Superiorization Methodology (SM) is to make use of the perturbations of (23) to transform a strongly perturbation resilient algorithm that seeks a constraints-compatible solution for C into one whose outputs are equally good from the point of view of constraints-compatibility, but are superior (not necessarily optimal) according to the objective function ϕ.

This is done by producing from the Basic Algorithm another algorithm, called its *superiorized* version, that makes sure not only that the $\beta_k v^k$ are bounded perturbations, but also that $\phi\left(y^k + \beta_k v^k\right) \leq \phi\left(y^k\right)$, for all $k \geq 0$. To do so, we use the next concept, closely related to the concept of "descent direction".

Definition 4 Given a function $\phi : R^J \to R$ and a point $y \in R^J$, we say that a vector $d \in R^J$ is `nonascending for` ϕ `at` y if $\|d\| \leq 1$ and there is a $\delta > 0$ such that

$$\text{for all } \lambda \in [0, \delta] \text{ we have } \phi(y + \lambda d) \leq \phi(y). \tag{24}$$

Obviously, the zero vector is always such a vector, but for superiorization to work we need a sharp inequality to occur in (24) frequently enough.

The Superiorized Version of the Basic Algorithm assumes that we have available a summable sequence $\{\eta_\ell\}_{\ell=0}^{\infty}$ of positive real numbers (for example, $\eta_\ell = a^\ell$, where $0 < a < 1$) and it generates, simultaneously with the sequence $\left\{y^k\right\}_{k=0}^{\infty}$ in Ω, sequences $\left\{v^k\right\}_{k=0}^{\infty}$ and $\{\beta_k\}_{k=0}^{\infty}$. The latter is generated as a subsequence of $\{\eta_\ell\}_{\ell=0}^{\infty}$, resulting in a nonnegative summable sequence $\{\beta_k\}_{k=0}^{\infty}$. The algorithm further depends on a specified initial point $y^0 \in \Omega$ and on a positive integer N. It makes use of a logical variable called *loop*. The superiorized algorithm is presented next by its pseudo-code.

Superiorized Version of the Basic Algorithm

set $k = 0$
set $y^k = y^0$
set $\ell = -1$
repeat
    set $n = 0$
    set $y^{k,n} = y^k$
    while $n<N$
        set $v^{k,n}$ to be a nonascending vector for ϕ at $y^{k,n}$
        set *loop*=*true*
        while *loop*
            set $\ell = \ell + 1$
            set $\beta_{k,n} = \eta_\ell$
            set $z = y^{k,n} + \beta_{k,n} v^{k,n}$
            if $\phi(z) \leq \phi\left(y^k\right)$ **then**
                set $n = n + 1$
                set $y^{k,n} = z$
                set *loop* = *false*
    set $y^{k+1} = \boldsymbol{A}_C\left(y^{k,N}\right)$

19. **set** $k = k + 1$

According to the analysis of the behavior of this algorithm in [22], the algorithm produces a sequence $\left\{y^k\right\}_{k=0}^{\infty}$ for which (23) is satisfied. Further, the following important fact is shown to be true in [22]: If, for a given $\varepsilon > 0$, the ε-output $O\left(C, \varepsilon, \left\{x^k\right\}_{k=0}^{\infty}\right)$ of the Basic Algorithm is defined for every $x^0 \in \Omega$, then every sequence $\left\{y^k\right\}_{k=0}^{\infty}$ generated by the Superiorized Version of the Basic Algorithm has an ε'-output $O\left(C, \varepsilon', \left\{y^k\right\}_{k=0}^{\infty}\right)$ for every $\varepsilon' > \varepsilon$. In other words, the Superiorized Version produces outputs that are essentially as constraints-compatible as those produced by the original (not superiorized) algorithm. However, due to the repeated steering of the process by lines 7 to 17 toward reducing the value of the objective function ϕ, we can expect that the output of the Superiorized Version will be superior (from the point of view of ϕ) to the output of the original algorithm.

5 A computational demonstration

5.1 The x-ray CT problem

The fully-discretized model in the series expansion approach to the image reconstruction problem of x-ray computerized tomography (CT) is formulated in the following manner. A Cartesian grid of square picture elements, called *pixels*, is introduced into the region of interest so that it covers the whole picture that has to be reconstructed. The pixels are numbered in some agreed manner, say from 1 (top left corner pixel) to J (bottom right corner pixel).

The x-ray attenuation function is assumed to take a constant value x_j throughout the jth pixel, for $j = 1, 2, ..., J$. Sources and detectors are assumed to be points and the rays between them are assumed to be lines. Further, assume that the length of intersection of the ith ray with the jth pixel, denoted by a_j^i, for $i = 1, 2, ..., I, \;\; j = 1, 2, ..., J$, represents the weight of the contribution of the jth pixel to the total attenuation along the ith ray.

The physical measurement of the total attenuation along the ith ray, denoted by b_i, represents the line integral of the unknown attenuation function along the path of the ray. Therefore, in this fully-discretized model, the line integral turns out to be a finite sum and the model is described by a system of linear equations

$$\sum_{j=1}^{J} x_j a_j^i = b_i, \quad i = 1, 2, \ldots, I. \tag{25}$$

In matrix notation we rewrite (25) as

$$Ax = b, \tag{26}$$

where $b \in R^I$ is the *measurement vector*, $x \in R^J$ is the *image vector*, and the $I \times J$ matrix $A = \left(a_j^i\right)$ is the *projection matrix*. See [20], especially Section 6.3, for a complete treatment of this subject.

5.2 The algorithms that we use

In this section we describe the PSM and SM algorithms specifically used in our demonstration. We applied both algorithms to solve the fully-discretized model in the series expansion approach to the image reconstruction problem of x-ray CT, formulated in the previous subsection and represented by the optimization problem

$$\text{minimize}\,\{\phi(x) \mid Ax = b \text{ and } 0 \le x \le 1\}\,. \tag{27}$$

The box constraints are natural for this problem: If x_j represents the linear attenuation coefficient, measured in cm^{-1}, at a medically-used x-ray energy spectrum in the jth pixel, then the box constraints $0 \le x \le 1$ are reasonable for tissues in the human body; see Table 4.1 of [20]. Hence, for the image reconstruction problem of x-ray CT, we define Ω by

$$\Omega = \left\{x \in R^J \mid 0 \le x \le 1\right\}. \tag{28}$$

We note that this Ω is bounded.

The choice of C in (1) is of the type specified in (3), with $L = I + 1$, $C_i = \left\{x \in R^J \mid \left\langle a^i, x\right\rangle = b_i\right\}$, for $i = 1, 2, \ldots, I$ and $C_{I+1} = \Omega$. Note that these are exactly the conditions that were used to derive in Section 3 the special case of the PSM expressed in (13) and the equations that follow it. Furthermore, since in the experiment reported below, we start with a specific image vector $x \in \Omega$ and calculate from it the measurement vector $b \in R^I$ using (25), we know that C is a nonempty subset of Ω, which is the requirement stated below (3).

For any such C, we define $Prox_C : \Omega \to R_+$ by

$$Prox_C(x) = \sqrt{\sum_{i=1}^{I} b_i - \langle a^i, x\rangle^2}. \tag{29}$$

Note that this proximity function $Prox_C$ is uniformly continuous and thus satisfies the condition stated for it in Theorem 1.

Our choice for the objective function ϕ is the total variation (TV) of the image vector x. Denoting the $G \times H$ image array X ($GH = J$) obtained from the image vector x by $X_{g,h} = x_{(g-1)H+h}$, for $1 \le g \le G$ and $1 \le h \le H$, we use

$$\phi\,(x) = \text{TV}(X) = \sum_{g=1}^{G-1} \sum_{h=1}^{H-1} \sqrt{\left(X_{g+1,h} - X_{g,h}\right)^2 + \left(X_{g,h+1} - X_{g,h}\right)^2}. \tag{30}$$

5.2.1 Projected Subgradient Method (PSM)

As mentioned at the beginning of Section 3, according to [27, Theorem 3.2.2], we need to verify convexity and local Lipschitz continuity of the objective function in (30). Convexity of the TV objective function follows, for example, from the end of the Proof of Proposition 1 in [16]. The Lipschitz continuity is not difficult to prove as the next proposition shows.

Proposition 1 *The total variation function (30) is Lipschitz continuous on the whole space R^J.*

Proof Each summand in (30) can be written as

$$\sqrt{\left(X_{g+1,h}-X_{g,h}\right)^2+\left(X_{g,h+1}-X_{g,h}\right)^2}=\left\|A_{g,h}X\right\|_2, \tag{31}$$

where $A_{g,h}$ is a square matrix having only two nonzero rows, with the first nonzero row containing only two nonzero elements 1 and -1 that correspond to the variables $X_{g+1,h}$ and $X_{g,h}$, respectively, and the second nonzero row containing only two nonzero elements 1 and -1 that correspond to the variables $X_{g,h+1}$ $X_{g,h}$, respectively. Thus, the TV function can be rewritten as

$$TV(X)=\sum_{g=1}^{G-1}\sum_{h=1}^{H-1}\left\|A_{g,h}X\right\|_2, \tag{32}$$

allowing us to estimate the Lipschitz constant as follows:

$$\begin{aligned}|TV(X)-TV(Y)| &= \left|\sum_{g=1}^{G-1}\sum_{h=1}^{H-1}\left\|A_{g,h}X\right\|_2-\left\|A_{g,h}Y\right\|_2\right| \\ \leq \sum_{g=1}^{G-1}\sum_{h=1}^{H-1}\left|\left\|A_{g,h}X\right\|_2-\left\|A_{g,h}Y\right\|_2\right| &\leq \sum_{g=1}^{G-1}\sum_{h=1}^{H-1}\left\|A_{g,h}X-A_{g,h}Y\right\|_2 \\ = \sum_{g=1}^{G-1}\sum_{h=1}^{H-1}\left\|A_{g,h}\left(X-Y\right)\right\|_2 &\leq \sum_{g=1}^{G-1}\sum_{h=1}^{H-1}\left\|A_{g,h}\right\|_2\left\|X-Y\right\|_2,\end{aligned}$$

where $\|A\|_2=\sqrt{\lambda_{\max}\left(A^TA\right)}$ with $\lambda_{\max}$ denoting the maximal eigenvalue. Thus we have shown

$$|TV(X)-TV(Y)|\leq L\left\|X-Y\right\|_2, \tag{33}$$

with $L=\sum_{g=1}^{G-1}\sum_{h=1}^{H-1}\left\|A_{g,h}\right\|_2$.

We note that the function $f(x)=\frac{1}{2}\left\|x-q^k\right\|^2$ is convex, continuously differentiable, and fulfills (17) with $L=1$. The algorithm that we use to realize the PSM is as follows.

The PSM Algorithm

(P1) **Initialization**: Select a point $x^0\in R^J$, select integers K and M, use two real number variables $\boldsymbol{curr}$ and $\boldsymbol{prev}$, and set $\boldsymbol{curr}=\phi\left(x^0\right)$ and $\boldsymbol{prev}=\boldsymbol{curr}$.

(P2) **Iterative step**: Given the current iterate x^k, calculate the next iterate x^{k+1} as follows:

(P2.1) Calculate a subgradient of ϕ at x^k, i.e., $\phi'\left(x^k\right)\in\partial\phi\left(x^k\right)$, a step-size $t_k=k^{-1/4}/\left\|\phi'\left(x^k\right)\right\|_2$ and the vector

$$q^k=x^k-t_k\phi'\left(x^k\right). \tag{34}$$

(P2.2) Apply Nesterov's Algorithm to solve the problem

$$\text{minimize}\left\{\frac{1}{2}\left\|x-q^k\right\|^2\mid Ax=b\text{ and }0\leq x\leq 1\right\}. \tag{35}$$

(P2.3) Update to find the next iterate

$$x^{k+1} = \arg\min_x \left\{ \frac{1}{2} \left\| x - q^k \right\|^2 \mid Ax = b \text{ and } 0 \le x \le 1 \right\}. \tag{36}$$

(P2.4) If $\phi\left(x^{k+1}\right) \le \boldsymbol{curr}$ then $\boldsymbol{curr} = \phi\left(x^{k+1}\right)$.
(P3) **Stopping rule**: If $k \, mod \, K = 0$ (i.e., k is divisible by K), then:
If $\boldsymbol{prev} - \boldsymbol{curr} < \boldsymbol{prev} / M$ then stop. Otherwise, $\boldsymbol{prev} = \boldsymbol{curr}$ and go to (2).

In step (P2.2) above, problem (35) is solved as follows. From a given problem (35) pass to its dual

$$\text{maximize} \left\{ f(\lambda) \mid \lambda \in R^I \right\}, \tag{37}$$

where

$$\begin{aligned} f(\lambda) = & \tfrac{1}{2} \left\| q^k - A^T\lambda - P_{C_{I+1}}\left(q^k - A^T\lambda\right) \right\|^2 - \tfrac{1}{2} \left\| q^k - A^T\lambda \right\|^2 \\ & - \langle \lambda, b \rangle + \tfrac{1}{2} \left\| q^k \right\|^2 \end{aligned} \tag{38}$$

and $P_{C_{I+1}}$ denotes the orthogonal projection onto the set $\left\{ x \in R^J \mid 0 \le x \le 1 \right\}$. Denote by λ^{*k} the optimal solution of problem (37). Then the optimal point x^{*k} of the problem (35) is given by

$$x^{*k} = P_{C_{I+1}}\left(q^k - A^T\lambda^{*k}\right). \tag{39}$$

We apply Nesterov's Algorithm for finding an optimal solution λ^{*k} of the problem (37).

In the reported experiments we used the starting points x^0 and y^0 in the PSM Algorithm and in Nesterov's Algorithm, respectively, to be zero vectors. In the initialization step of the PSM Algorithm we selected $K = 10$ and $M = 5000$. In the initialization step of Nesterov's Algorithm we chose $\alpha_{-1} = 10$.

5.2.2 Superiorization Method (SM)

Our selected choice for the operator $\boldsymbol{A}_C$ in the Basic Algorithm as well as in the Superiorized Version of the Basic Algorithm, as described in Section 4, is based on an algebraic reconstruction technique (ART), see [20, Chapter 11]. Specifically, for $i = 1, 2, \ldots, I$, we define the operators $U_i : R^J \to R^J$ by

$$U_i(x) = x + \frac{b_i - \left\langle a^i, x \right\rangle}{\left\| a^i \right\|^2} a^i. \tag{40}$$

Defining $Q : R^J \to \Omega$ by

$$\left(Q(x)\right)_j = \begin{cases} x_j, \text{ if } 0 \le x_j \le 1, \\ 0, \quad \text{if } x_j < 0, \\ 1, \quad \text{if } 1 < x_j, \end{cases} \tag{41}$$

for $j = 1, 2, ..., J$, we specify the algorithmic operator $\boldsymbol{A}_C : \Omega \to \Omega$ by

$$\boldsymbol{A}_C(x) = QU_I \cdots U_2 U_1(x). \tag{42}$$

Since the individual U_is as well as the Q are clearly nonexpansive operators, the same is true for $\boldsymbol{A}_C$.

By well-known properties of ART (see, for example, Sections 11.2 and 15.8 of [20]), the Basic Algorithm with this algorithmic operator is convergent over Ω and, in fact, for every $x^0 \in \Omega$, the limit $y\left(x^0\right)$ is in C. It follows that, for every $x^0 \in \Omega$, $Prox_C\left(y\left(x^0\right)\right) = 0$ and so the Basic Algorithm is boundedly convergent. According to Theorem 1, this combined with the facts that $\boldsymbol{A}_C$ is nonexpansive and the proximity function $Prox_C$ is uniformly continuous, implies that the Basic Algorithm defined by $\boldsymbol{A}_C$ is strongly perturbation resilient.

Recalling the discussion just below the description of the Superiorized Version of the Basic Algorithm, we make further use of the convergence of the Basic algorithm to an element of C as follows. Since for all $\varepsilon > 0$, the ε-output $O\left(C, \varepsilon, \left\{x^k\right\}_{k=0}^{\infty}\right)$ of the Basic Algorithm is defined for every $x^0 \in \Omega$, we also have that every sequence $\left\{y^k\right\}_{k=0}^{\infty}$ generated by the Superiorized Version of the Basic Algorithm has an ε'-output $O\left(C, \varepsilon', \left\{y^k\right\}_{k=0}^{\infty}\right)$ for every $\varepsilon' > 0$. This means that for the specific type of C that is used in our comparative study, the Superiorized Version of the Basic Algorithm is guaranteed to produce an ε'-compatible output for any $\varepsilon' > 0$ and any initial point $y^0 \in \Omega$.

The specific choices made when running the Superiorized Version of the Basic Algorithm for our comparative study were the following. We selected $\eta_\ell = 0.999^\ell$, y^0 to be the zero vector and $N = 9$. All these choices we made are based on auxiliary experiments (not included in this paper) that helped determine optimal parameters for the data-set discussed in Subsection 5.3. In addition, we need to specify how the nonascending vector was selected in line 8 of the Superiorized Version of the Basic algorithm. This was done by the method specified in [22], Section III.A following equation (12).

5.3 The computational result

The computational work reported here was done on a single machine using a single CPU, an Intel i5-3570K 3.4 Ghz with 16 GB RAM using the SNARK09 software package [18,24]; the phantom, the data, the reconstructions and displays were all generated within this same framework. In particular, this implies that differences in the reported reconstruction times are not due to the different algorithms being implemented in different environments.

Figure 1 shows the phantom used in our study, which is a 485×485 digitized image. The phantom corresponds to a cross-section of a human head (based on [20, Figure 4.6]). It is represented by a vector with $235,225$ components, each standing for the average x-ray attenuation coefficient within a pixel. Each pixel is of size 0.376×0.376 mm^2. The values of the components are in the range of [0, 0.6241749], however the display range used here was much smaller [0.204, 0.21675]. The mapping between the two ranges is such that any value below 0.204 is shown as black and any value above 0.21675 is shown as white with a linear mapping in-between. We used this display window for all images presented here.

Data were collected by calculating line integrals through the digitized head phantom in Figure 1 using 60 sets of equally rotated (in 3 degrees increments) parallel lines, with

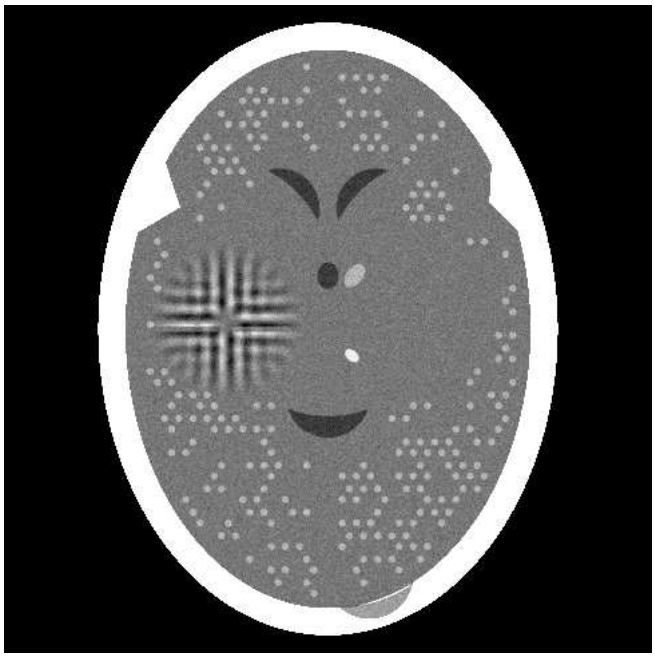


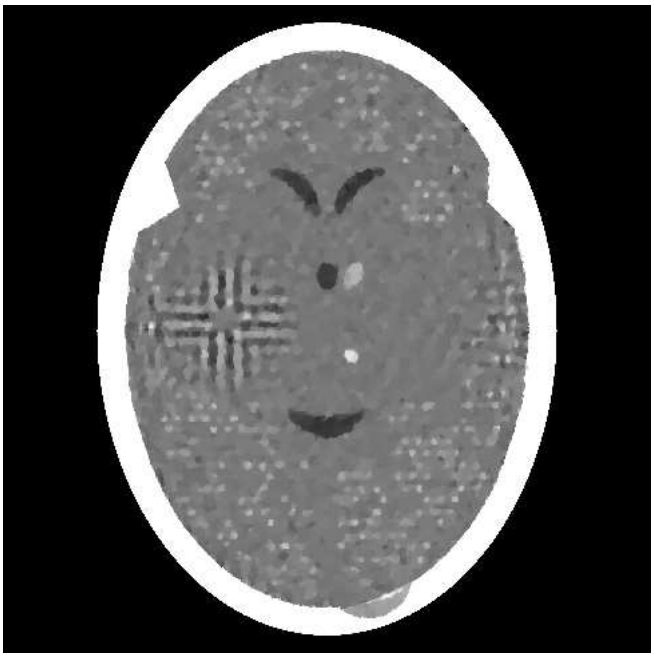
Fig. 1: The head phantom. The value of its TV is 984. Its tomographic data was obtained for 60 views.

lines in each set spaced at 0.752 mm from each other. Each line integral gives rise to a linear equation and represents a hyperplane in R^J. The phantom itself lies in the intersection of all the hyperplanes that are associated with these lines and it also satisfies the box constraints in (28) The total number of linear equations is $18,524$, making our problem underdetermined with $235,225$ unknowns (the intersection of all the hyperplanes is an at least in a $216,701$-dimensional subspace of $R^{235,225}$). In the comparative study, we first applied the PSM and then the SM to these data as follows.

The PSM was implemented as described in Subsection 5.2.1. In particular, it started with $x^0 = 0$ (the zero vector all of whose components are 0), for which $Prox_C\left(x^0\right) = 326$. It was stopped according to the Stopping Rule (P3), with $K = 10$ and $M = 5000$. The iteration number at that time was 815 and the value of the proximity function was $Prox_C\left(x^{815}\right) = 0.0422$, which is very much smaller than the value at the initial point. The length of computer time required was 2217 seconds. The TV of the output was 919, which is less than that of the phantom, indicating that PSM is performing its task of producing a constraints-compatible output with a low TV. This output is shown in Figure 2(a).

We used the Superiorized Version of the Basic Algorithm, as described in Subsection 5.2.2 to generate a sequence $\left\{y^k\right\}_{k=0}^{\infty}$ until it reached $O\left(C, 0.0422, \left\{y^k\right\}_{k=0}^{\infty}\right)$ and considered that to be the output of SM. We know that this output must exist for our problem and that its constraints-compatibility will not be greater than that of the output of PSM. The length of computer time required to obtain this output was 102 seconds, which is over twenty times shorter than what was needed by PSM to get its

(a)



(b)

Fig. 2: Reconstructions of the head phantom of Figure 1. (a) The image reconstructed by the PSM has $TV = 919$ and was obtained after 2217 seconds. (b) The image reconstructed by the SM has $TV = 873$ and was obtained after 102 seconds.

	TV value	Time (seconds)
PSM	919	2217
SM	873	102

Table 1: Performance comparison of PSM and SM when producing the reconstructions in Figure 2.

output. The TV of the SM output was 876, which is also less than that of the output of PSM. The SM output is shown in Figure 2(b).

As summarized in Table 1, with the stopping rule that guarantees that the output of the SM is at least as constraints-compatible as the output of the PSM, the SM showed clearly superior efficacy to the PSM: it obtained a result with a lower TV value at less than one twentieth of the computational cost.

6 Conclusions

The superiorization methodology (SM) allows the conversion of a feasibility-seeking algorithm, designed to find an ε-compatible solution of the constraints, into a superiorized algorithm that inserts, into the feasibility-seeking algorithm, objective function reduction steps without ruining the guaranteed feasibility-seeking nature of the algorithm. The superiorized algorithm interlaces objective function nonascent steps into the original process in an automatic manner. In case of strong perturbation resilience of the original feasibility-seeking algorithm, mathematical results indicate why the superiorized algorithm will be efficacious for producing an an ε-compatible solution output with a low value of the objective function.

We have presented an example for which the SM finds a better solution to a constrained minimization problems than the projected subgradient method (PSM), and in significantly less computation time. This finding is understandable in view of the nature of how the methods interlace feasibility oriented activities with optimization activities. While the PSM requires a projection onto the feasible region of the constrained minimization problem, the SM needs to do only projections onto the individual constraints whose intersection is the feasible region. We demonstrated this experimentally on a large-sized application that is modeled, and needs to be solved, as a constrained minimization problem.

Acknowledgments. We would like to acknowledge the generous support by Dr. Ernesto Gomez and Dr. Keith Schubert in allowing us to use the GP-GPU cluster at the Department of Computer Science and Engineering at California State University San Bernardino. We are also grateful to Joanna Klukowska for her advice on using optimized compilation for speeding up SNARK09. This work was supported by the United States-Israel Binational Science Foundation (BSF) Grant No. 200912, the U.S. Department of Defense Prostate Cancer Research Program Award No. W81XWH-12-1-0122, the National Science Foundation Award No. DMS-1114901, the U.S. Department of Army Award No. W81XWH-10-1-0170, and by Grant No. R01EB013118 from the National Institute of Biomedical Imaging and Bioengineering and the National Science Foundation. The contents of this publication is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Biomedical Imaging and Bioengineering or the National Institutes of Health.

References

1. Aharoni, R., Censor, Y.: Block-iterative projection methods for parallel computation of solutions to convex feasibility problems. Linear Algebra Appl. **120**, 165–175 (1989)
2. Bauschke, H.H., Borwein, J.M.: On projection algorithms for solving convex feasibility problems. SIAM Rev. **38**, 367–426 (1996)
3. Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer, New York, NY, USA (2011)
4. Bauschke, H.H., Koch, V.R.: Projection methods: Swiss army knives for solving feasibility and best approximation problems with halfspaces. Tech. rep., http://arxiv.org/abs/1301.4506v1 (2012)
5. Beck, A., Teboulle, M.: Mirror descent and nonlinear projected subgradient methods for convex optimization. Oper. Res. Lett. **31**, 167–175 (2003)
6. Boyd, S., Mutapcic, A.: Subgradient methods (2007). Course Notes, *http* : *//www.stanford.edu/class/ee*364*b/notes/subgrad_method_notes.pdf*
7. Butnariu, D., Davidi, R., Herman, G.T., Kazantsev, I.G.: Stable convergence behavior under summable perturbations of a class of projection methods for convex feasibility and optimization problems. IEEE J. Sel. Topics Signal Process. **1**, 540–547 (2007)
8. Cegielski, A.: Iterative methods for Fixed Point Problems in Hilbert Spaces. Lecture Notes in Mathematics 2057, Springer-Verlag, Berlin, Heidelberg, Germany (2012)
9. Censor, Y., Chen, W., Combettes, P.L., Davidi, R., Herman, G.T.: On the effectiveness of projection methods for convex feasibility problems with linear inequality constraints. Comput. Optim. Appl. **51**, 1065–1088 (2012)
10. Censor, Y., Davidi, R., Herman, G.T.: Perturbation resilience and superiorization of iterative algorithms. Inverse Problems **26**, 065,008 (2010)
11. Censor, Y., Elfving, T., Herman, G.T.: Averaging strings of sequential iterations for convex feasibility problems. In: D. Butnariu, Y. Censor, S. Reich (eds.) Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications, pp. 101–114. Elsevier Science Publishers, Amsterdam (2001)
12. Censor, Y., Segal, A.: On the string averaging method for sparse common fixed point problems. Int. Trans. Oper. Res. **16**, 481–494 (2009)
13. Censor, Y., Segal, A.: On string-averaging for sparse problems and on the split common fixed point problem. Contemp. Math. **513**, 125–142 (2010)
14. Censor, Y., Tom, E.: Convergence of string-averaging projection schemes for inconsistent convex feasibility problems. Optim. Methods Softw. **18**, 543–554 (2003)
15. Censor, Y., Zenios, S.A.: Parallel Optimization: Theory, Algorithms, and Applications. Oxford University Press, New York, NY, USA (1997)
16. Combettes, P.L., Pesquet, J.C.: Image restoration subject to a total variation constraint. IEEE Trans. Image Process. **13**, 1213–1222 (2004)
17. Davidi, R., Herman, G.T., Censor, Y.: Perturbation-resilient block-iterative projection methods with application to image reconstruction from projections. Int. Trans. Oper. Res. **16**, 505–524 (2009)
18. Davidi, R., Herman, G.T., Klukowska, J.: SNARK09: A programming system for the reconstruction of 2D images from 1D projections. http://www.dig.cs.gc.cuny.edu/software/snark09/ (2009)
19. Fletcher, R., Leyffer, S.: Nonlinear programming without a penalty function. Math. Program. Ser. A **91**, 239–269 (2002)
20. Herman, G.T.: Fundamentals of Computerized Tomography: Image Reconstruction from Projections, 2nd edn. Springer (2009)
21. Herman, G.T., Davidi, R.: Image reconstruction from a small number of projections. Inverse Problems **24**, 045,011 (2008)
22. Herman, G.T., Garduño, E., Davidi, R., Censor, Y.: Superiorization: An optimization heuristic for medical physics. Med. Phys. **39**, 5532–5546 (2012)
23. Jin, W., Censor, Y., Jiang, M.: A heuristic superiorization-like approach to bioluminescence tomography. In: International Federation for Medical and Biological Engineering (IFMBE) Proceedings, vol. 39, pp. 1026–1029 (2012)
24. Klukowska, J., Davidi, R., Herman, G.T.: SNARK09 - A software package for reconstruction of 2D images from 1D projections. Comput. Methods Programs Biomed. (to appear)
25. Luo, S., Zhou, T.: Superiorization of em algorithm and its application in single-photon emission computed tomography (spect). Tech. rep., http://arxiv.org/abs/1209.6116 (2012)

26. Nedić, A.: Random algorithms for convex minimization problems. Math. Program. Ser. B **129**, 225–253 (2011)
27. Nesterov, Y.: Introductory Lectures on Convex Optimization. Kluwer Academic Publishers, Boston/Doredrecht/London (2004)
28. Nesterov, Y.E.: A method for solving the convex programming problem with convergence rate $O(1/k^2)$. Soviet Math. Doklady **27**, 372–376 (1983)
29. Neto, E.S.H., De Pierro, Á.R.: On perturbed steepest descent methods with inexact line search for bilevel convex optimization. Optimization **60**, 991–1008 (2011)
30. Neto, E.S.H., Pierro, Á.R.D.: Incremental subgradients for constrained convex optimization: A unified framework and new methods. SIAM J. Optim. **20**, 1547–1572 (2009)
31. Nikazad, T., Davidi, R., Herman, G.: Accelerated perturbation-resilient block-iterative projection methods with application to image reconstruction. Inverse Problems **28**, 035,005 (2012)
32. Nurminski, E.: The use of additional diminishing disturbances in Fejer models of iterative algorithms. Comput. Math. Math. Phys. **48**, 2154–2161 (2008). Original Russian Text: E.A. Nurminski, published in: Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki 48 (2008), 2121–2128
33. Nurminski, E.A.: Fejer processes with diminishing disturbances. Doklady Mathematics **78**, 755–758 (2008). Original Russian text: E.A. Nurminski, published in: Doklady Akademii Nauk 422 (2008), 601–604
34. Nurminski, E.A.: Envelope stepsize control for iterative algorithms based on Fejer processes with attractants. Optim. Methods Softw. **25**, 97–108 (2010)
35. Nurminski, E.A.: Fejer algorithms with an adaptive step. Comput. Math. Math. Phys. **51**, 741–750 (2011). Original Russian text: E.A. Nurminski, published in: Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki 51 (2011), 791–801
36. Penfold, S.N., Schulte, R.W., Censor, Y., Bashkirov, V., McAllister, S., Schubert, K.E., Rosenfeld, A.B.: Block-iterative and string-averaging projection algorithms in proton computed tomography image reconstruction. In: Y. Censor, M. Jiang, G. Wang (eds.) Biomedical Mathematics: Promising Directions in Imaging, Therapy Planning and Inverse Problems, pp. 347–367. Medical Physics Publishing, Madison, WI, USA (2010)
37. Penfold, S.N., Schulte, R.W., Censor, Y., Rosenfeld, A.B.: Total variation superiorization schemes in proton computed tomography image reconstruction. Med. Phys. **37**, 5887–5895 (2010)
38. Poljak, B.T.: A general method of solving extremum problems. Soviet Math. Dokl. **8**, 593–597 (1967)
39. Polyak, B.T.: Minimization of unsmooth functionals. USSR Comput. Maths. Math. Phys. **9**, 14–29 (1969)
40. Ram, S.S., Nedić, A., Veeravalli, V.: Incremental stochastic subgradient algorithms for convex optimization. SIAM J. Optim. **20**, 691–717 (2009)
41. Ruszczyński, A.: Nonlinear Optimization. Princeton University Press, Princeton, NJ, USA (2006)
42. Shor, N.Z.: Minimization Methods for Non-Differentiable Functions. Springer-Verlag, Berlin, Heidelberg, Germany (1985)
43. Sidky, E.Y., Duchin, Y., Pan, X., Ullberg, C.: A constrained, total-variation minimization algorithm for low-intensity x-ray CT. Med. Phys. **38**, S117–S125 (2011)
44. Sidky, E.Y., Kao, C., Pan, X.: Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. Phys. Med. Biol. **53**, 4777–4807 (2008)